HOCHSCHULE
MITTWEIDA
UNIVERSITY OF
APPLIED SCIENCES

BACHELORARBEIT

Herr
Christian Klaus

Lösungsansätze der IT-Forensik zur Erkennung kinder- und jugend-pornografischer Inhalte

Mittweida, 2017

Fakultät Angewandte Computer- und Biowissenschaften

BACHELORARBEIT

Lösungsansätze der IT-Forensik zur Erkennung kinder- und jugend-pornografischer Inhalte

Autor:
Herr

Christian Klaus

Studiengang:
Allgemeine und Digitale Forensik

Seminargruppe:
FO14w2-B

Erstprüfer:
Prof. Dr. rer. nat. Christian Hummert

Zweitprüfer:
Dipl. Ing. Mag. Marian Kogler

Einreichung:
Mittweida, 11.08.2017

Verteidigung/Bewertung:
Mittweida, 2017

Faculty Applied Computer Science and Biosciences

BACHELORTHESIS

IT-forensics solution approaches in detecting child and teen pornography

author:
Mr.

Christian Klaus

course of studies:
General and Digital Forensic Science

seminar group:
FO14w2-B

first examiner:
Prof. Dr. rer. nat. Christian Hummert

second examiner:
Dipl. Ing. Mag. Marian Kogler

submission:
Mittweida, 11.08.2017

defence/ evaluation:
Mittweida, 2017

Bibliografische Beschreibung:

Klaus, Christian:
Lösungsansätze der IT-Forensik zur Erkennung kinder- und jugendpornografischer Inhalte. - 2017. - XI, 69, XXI S.
Mittweida, Hochschule Mittweida, Fakultät Angewandte Computer- und Biowissenschaften, Bachelorarbeit, 2017

Referat:

Die vorliegende Arbeit befasst sich mit der Analyse von Videos mit Kinder-und Jugendschauspielern, die durch die Forensik-Analyse-Software X-Ways Forensics gecarvt und gefiltert und von dem Gesichtserkennungsprogramm SHORE des Fraunhofer Instituts für Integrierte Schaltungen bezüglich Alter uns Geschlecht analysiert werden. Die Biometrische Unterscheidung des Alters ist besonders bei der Erkennung und Differenzierung zwischen legaler Erwachsenenpornographie und inkriminierter Kinder- sowie Jugendpornografie entscheidend. Die Ergebnisse werden unter verfahrensechten Umständen getestet und auf Gerichtsverwertbarkeit geprüft. Um Vergleiche mit ähnlichen Produkten und der manuellen Arbeit eines erfahrenen Ermittlers zu erstellen, werden Testdaten mit bekannten Parametern eingespeist und bezüglich Treffer- und Fehlergenauigkeit sowie Dauer verglichen.

Abstract:

In order to identifying incriminated data and dividing it from legal adult pornography, this thesis provides testing videos and images for the face recognition software SHORE from the Fraunhofer Institute IIS Berlin, a programme fully capable of biometric face recognition, gender and emotion analysis and age estimation, making it a valuable and crucial tool for child and teen pornography detection and distinction. Furthermore the test-set of videos with known parameters is also provided to similar software solutions as well as the manual work of an experienced forensic analyst. They are compared regarding precision, recall and duration. This thesis also examines neural networks and pornography scanners, testing their capabilities and errors.

Inhalt

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

Abb.	Abbildung
API	Application Programming Interface (Programmierschnittstelle)
CCA	Canonical Correlation Analysis (kanonische Korrelationsanalyse)
CCTV	Close Circuit Television (Videoüberwachungsanlage)
etc.	et cetera (und die übrigen)
GF	Gabor Filter
GUI	Graphical User Interface (Grafische Oberfläche)
HOG	Histogram of oriented gradients
ICA	Independent Component Analysis
KCCA	Kernel Canonical Correlation Analysis
KPCA	Kernel Principal Component Analysis
LBP	Local Binary Pattern (Lokales Binärmuster)
LDA	Linear Discriminant Analysis (Lineare Diskriminanzanalyse)
Männl.	Männlich
PCA	Principal Component Analysis (Hauptkomponentenanalyse)
PLS	Partial Least Squares
S.	Seite
SDK	Software Development Kit
Weibl.	Weiblich
XWF	X-Ways Forensics

1 Übersicht

1.1 Motivation

Die digitale Erkennung eines menschlichen Gesichts und der Abgleich mit Vergleichsmaterial sind zwei Kernthemen, die für die Wissenschaft lange Zeit eine Problemquelle dargestellt haben. Allerdings haben sich in den letzten Jahren diesbezüglich einige enorme Entwicklungen durchgesetzt. Immer modernere Methoden und Tools ermöglichten 3D-Rekonstruktion aus Videomaterial, biometrische Bild- und Videoanalysen mit Gesichtsvergleichen und automatische Gesichtserkennung aus CCTV-Aufnahmen. Doch noch immer stellt die genaue Alters- und Geschlechtsbestimmung eine Herausforderung für die Forensik dar. Gerade dabei handelt es sich um ein entscheidendes Merkmal vieler Ermittlungen, insbesondere im Bereich der Kinder- und Jugendpornografie. Denn eine eindeutige Unterscheidung an diversen Altersgrenzen, basierend auf genauen Algorithmen und nicht auf dem reinen Abschätzen eines mehr oder weniger erfahrenen Ermittlers, könnte zum zweifelsfreien, gerichtsfesten Beweis für eine Straftat werden.

Zum jetzigen Zeitpunkt sieht die Realität jedoch anders aus: hunderte Ermittler privater Unternehmen und Regierungsbehörden weltweit arbeiten sich täglich manuell durch Berge eventuell inkriminierter Daten, vor allem Foto und Videodateien. Jede Einzelne muss per Hand hinsichtlich ihres Inhaltes gesichtet werden. Dieser Vorgang ist sowohl zeit- als auch ressourcenaufwändig, trotz langjähriger Erfahrung nur bedingt genau und für die betreffenden Ermittler psychisch belastend. Die Realisierung eines automatischen Systems wurde unter anderem von Videoanalysesoftware wie SHORE und Ähnlichen implementiert; Programme, die über die Algorithmen verfügen, Gesichter in Videos zu erkennen und Alter und Geschlecht zu bestimmen.

1.2 Zielsetzung

Ziel dieser Arbeit ist es zum Einen einen Überblick über verschiedene, modernste Methoden zur automatischen Erkennung von inkriminierten Daten zu erarbeiten und einen speziell ausgewählten Datensatz zu erstellen.

Sämtliche zu untersuchende Dateien werden von der Forensik-Suite X-Ways Forensics exportiert und an die Analysesoftware des Fraunhofer Instituts für Integrierte Schaltungen SHORE übergeben und dort verarbeitet.

Zum Anderen soll ein Vergleich bezüglich der Genauigkeit und dem Zeitaufwand zwischen SHORE und ähnlichen Erkennungssoftwares und ebenso einem menschlichen Ermittler gezogen werden. Anschließend werden weitere Vorgehensweisen zur Erkennung von kinder- und jugendpornografischen Inhalten vorgestellt und getestet.

1.3 Kapitelüberblick

Das folgende Kapitel 2 gilt als Einleitung in die Materie, hier werden die benutzten Programme wie X-Ways Forensics und SHORE erörtert, sowie die grundsätzlichen Funktionen und Vorgehensweisen von Biometrie und Gesichtserkennung erklärt, aufgrund denen SHORE, OpenFace, Kairos und andere Programme funktionieren. Das Kapitel stellt vorrangig genutzte Algorithmen vor, die in den Videodateien Gesichter erkennen und sowohl das Alter als auch das Geschlecht bestimmen.

Kapitel 3 umfasst die einzelnen Schritte der Videoauswahl, der Videoextraktion aus X-Ways Forensics und der Analyse durch SHORE.

Kapitel 4 beschreibt die Ergebnisse des Projektes, das heißt, welche Resultate SHORE im Betracht auf Zeitaufwand und Erkennungsgenauigkeit vorweist . Hierbei werden Alters- und Geschlechtsbestimmung aller drei Altersgruppen untersucht und Differenzen und Abhängigkeiten erörtert.

Das daraufhin folgende Kapitel 5 vergleicht die in Kapitel 4 vorgestellten Ergebnisse mit denen ähnlicher Produkte und der Arbeitsweise eines Ermittlers, der die Daten händisch auswerten muss. Außerdem liefert das Kapitel einen vorausschauenden Ausblick auf weitere Möglichkeiten und Verbesserungen sowie ein zusammenfassendes Fazit.

.

2 Einleitung

2.1 X-Ways Forensics

Im Jahr 2002 gründete der Deutsche IT-Sicherheits-Experte Stefan Fleischmann die Firma X-Ways Software Technologie AG, die bis zum Sommer 2004 zwei separat arbeitende Programme veröffentlichte, den "Hex- und Disk-Editor Winhex" und das forensische Analyse-Tool XWF, "X-Ways Forensics" [1 S. XVIII].

Der Name X-Ways stammt laut Gründer daher, dass das Forensik-Programm so umfassend ist, dass es alle möglichen Probleme und Fragestellungen auf x-verschiedene Arten und Weisen lösen kann, also x Wege vom Programmstart zur Problemlösung beherrscht [1 S. XVIII.

Winhex ist ein universeller Hexadezimal-Editor, der besondere Features für die "forensische Untersuchung von EDV-Systemen, für Datenrettung und IT-Sicherheit"[1] aufweist. Dateisysteme der FAT-Familie, NTFS, Ext3/4 und Weitere können editiert, zerlegt und verkettet werden, Datenträger geklont, Zahlenformate konvertiert und vertrauliche Daten gelöscht und verschlüsselt werden.

Dahingegen ist die Forensik-Suite X-Ways Forensics "eine hochintegrierte Arbeitsumgebung für Computerspezialisten bei der forensischen (kriminaltechnischen) Untersuchung von EDV"[2], die mit umfangreichen Funktionen ausgestattet ist. Neben der Analyse von Datenträgern aller Art, verschiedensten Filtermethoden, der logischen, physikalischen und byteweisen Suche nach bestimmten Dateien oder Dateifragmenten, dem Auslesen von relevanten Systemkomponenten und dem Wiederherstellen von gelöschten Daten mittels File-Carving, kann das Programm auch getarnte Dateien per Dateiheader-Signatur-Suche überprüfen.

---

[1] X-Ways AG, "WinHex: Software für Computerforensik und Datenrettung, Hex-Editor und Disk-Editor", https://www.x-ways.net/winhex/index-d.html, angesehen am 27. April 2017

[2] X-Ways AG, "X-Ways Forensics: Integrierte Software für Computerforensik", https://www.x-ways.net/forensics/index-d.html, angesehen am 27. April 2017

Im Gegensatz zu internationalen Konkurrenten ist XWF trotz seiner Funktionsvielfalt sehr klein (nur 30 Megabytes in Version 18.5) und vollständig portabel, das heißt es kann ohne Installation oder Anlegen von Datenbanken beispielweise von einem USB-Stick ausgeführt werden. Ebenso ist es weniger ressourcenaufwändig und kompromittierbar wie ähnliche Forensik-Lösungen [1 S. XV].

Grundlegend verwendet XWF einen separierten Fall-Aufbau, das heißt, die zu bearbeitenden Daten(-träger) werden einem beschrifteten Fall untergeordnet, der mit Aktenzeichen, Fallnummern, Bearbeiter und Bearbeitungszeitraum ausgewiesen ist. Somit bleiben sämtliche Dateien dem Fall zugehörend, auch wenn temporäre Dateien, Unterasservate oder Exporte hinzukommen.

XWF erlaubt es Daten als inkriminiert zu markieren und Berichten hinzuzufügen, interne Kommentare zu Dateien zu verfassen und gesicherte Container zu erstellen. Es werden diverse Möglichkeiten zur Ansicht von Daten geboten, darunter eine hexadezimale (editorartige) Ansicht, Thumbnail-Galerien und die benutzerangepasste Integration von externen Viewer-Programmen zum Öffnen von Dokumenten und E-Mails, Video-, Audio- und Fotodateien, diversen Archiven sowie weiteren Formaten.

Staatsanwaltschaften und Ermittlungsbehörden greifen auf das Produkt zurück, weil es durch seinen vollständigen Schreibschutz verhindert, dass Daten vom Ermittler verändert werden und die implementierten Hash-Verfahren einen eindeutigen, gerichtsfesten Beweis für die Datenintegrität darstellen. Dabei werden sowohl vor als auch nach der Analyse der Asservate Hash-Werte, also eindeutige 128 oder 256 Bit große Darstellungen der Daten mit MD5 (Message Digest Algorithm 5) und SHA256 (Secure Hash Algorithm) erzeugt, die miteinander abgeglichen werden können und somit bei exakter Übereinstimmung beweisen, dass der bearbeitende Gutachter oder Ermittler selbst keinerlei Daten während des Analyseprozesses verändert hat. Die Sicherstellung dieses Vorgehens ist für gerichtsfeste Gutachten zwingend erforderlich.

Des Weiteren kann XWF natürlich auch Hashes für beliebige Daten erzeugen und verfügt ebenfalls über mehrere Hash-Datenbanken für Dateien, Bilder und Dokumente, die für das Black- und Whitelisting genutzt werden können, also den Abgleich mit zum Beispiel schon

bekannten Schadprogrammen oder inkriminierten Bildern respektive Dateien, die von anderen offiziellen Stellen als saubere, ungefährliche Dateien bestätigt wurden. Hierzu dienen unter anderem die umfangreichen Reference Data Sets (RDS) der National Software Reference Library (NSRL).

Bis heute findet die umfassende Lösung für Datenanalyse und -wiederherstellung mit weltweit über 35.000 Benutzern Verwendung in Ermittlungsbehörden, Staatsanwaltschaften und Unternehmen mit Hintergrund in der IT-Forensik und IT-Sicherheit [1 S. XVIII].

Die Verwendung von X-Ways Forensics ist für den Versuch dieser Arbeit nicht zwingend notwendig, dennoch wird auf das Tool zurückgegriffen, da der genaue Ablauf der Analyse eines Echtfalles simuliert werden soll. Hierbei ist die Untersuchung und Aufbereitung der Daten durch XWF immer gegeben.

2.2 Gesichtserkennung

Die automatische Gesichtserkennung stellt eine immer größer werdende Thematik dar, die im Alltagsleben des 21. Jahrhunderts eine tragende Rolle spielt, eine Zeit in der biometrische Authentifizierung, moderne Überwachungssysteme sowie Gesichtserkennung für Multimediaprojekte an der Tagesordnung stehen. Neben 3D-Designern der Film- und Videospielbranche und Fans von Facebooks Live Video Masks haben weltweit auch Sicherheitsexperten und Ermittlungsbehörden ein großes Interesse an der automatischen Erkennung eines Gesichtes und der eindeutigen Zuordnung zu einer Person entwickelt.

Dies beruht darauf, dass ein entscheidender Vorteil gegenüber Iris- oder Fingerabdruck-erkennung darin besteht, dass Gesichter auch über größere Entfernungen schnell und einfach erfasst werden können, dabei dennoch bei allen Menschen hoch divergent vorhanden sind und sich über längere Zeitabstände wenig verändern [2 S. 1].

Während es dem Menschen gemeinhin recht leicht fällt Gesichter zu erkennen und auch voneinander zu unterscheiden, ist es für die Computersoftware schwierig, denn zum reinen Feststellen des Vorhandenseins eines Gesichtes erfordert es nicht nur einfaches "Sehen", sondern unzählige Gleichungen, Variablen und Algorithmen.

Zum Einen wird hierbei gefordert, dass das System Gesichter unabhängig von dem Lichtverhältnis und -einfall, der Blickrichtung des Subjektes und seiner Emotionen ein Gesicht eindeutig identifizieren kann [2 S. 2] [7 S. XI]. Dabei muss die zugrundeliegende Algorithmik so ausgearbeitet sein, dass jedes Gesicht erkannt wird, und nichts erkannt wird, dass gar kein Gesicht ist.

Zum Anderen muss sich dieses Gleichungssystem aber als so flexibel herausstellen, dass es in der Lage ist Gesichter unterschiedlicher Personen mit unterschiedlichen Geschlechtern, Hautfarben, Altern und Emotionen erkennt. Ist dieses Verfahren erfolgreich implementiert, muss anschließend dafür Sorge getragen werden, dass selbst bei minderoptimalen Bedingungen die Erkennungsrate möglichst hoch bleibt [2 S 2, 7] [3].

Seit dem Beginn der Entwicklung von automatisierten Gesichtserkennungsalgorithmen in den 1960er Jahren, stellt die Frage nach der richtigen Gleichung die Wissenschaft vor eine große Herausforderung, eine, die bis heute nicht vollumfassend gelöst wurde.

Denn trotz modernster Technologie, enorm leistungsfähigen Supercomputern und ausgefeilten Mustervorlagen beweisen viele hochfunktionale und ebenso kostenintensive Produkte und Projekte zur Gesichtserkennung ihre Makel durch (zu) hohe Fehlerquoten, ungenaue Schätzungen und starke Vorgabenabweichung. Inzwischen sind einige topmoderne Systeme bereits so optimiert, dass sie die Fähigkeit zur Gesichtserkennung und Unterscheidung so effizient beherrschen, dass sie diesbezüglich den Menschen übertreffen können. Dies erfordert jedoch eine strenge Kontrolle der Umgebungssituation, das heißt alle Bilder müssen bestimmten Grundvorrausetzungen entsprechen, die als ideal betrachtet werden und die Gesichtserkennung begünstigen [2 S. 3] [3] [7 S. XI-XII].

Als Beispiele wären hier Passbilder oder Fahndungsfotos zu nennen, auf denen das unbewegte Gesicht möglichst emotionslos mit geschlossenem Mund, offenen Augen und ruhiger Stirn gerade in die Kamera schaut. Auch die Belichtung wird hier speziell angepasst. Versuche, die ausschließlich mit Bildern unter diesen Bedingungen stattfinden, werden auch kooperative Benutzerszenarios genannt, weil der Anwender hier das primäre Anliegen hat, dass sein Gesicht erkannt wird, zum Beispiel bei einem biometrischen Sicherheitsschloss oder einer E-Passport-Identifikation [2 S. 3].

Besondere Schwierigkeiten haben die Erkennungssysteme jedoch mit dem Arbeiten in sogenannten "unconstrained environments" oder auch "noncooperative user scenarios", also uneingeschränkten Umgebungen oder nicht-kooperativen Benutzerszenarios.

Diese treten auf, wenn die Rahmenbedingungen eines Vergleiches stark gelockert sind und damit an echte Lebensbedingungen angepasst wurden.

In diesen Fällen muss das Erkennungssystem mit unvorbereiteten Bildern arbeiten, bei denen die Chancen auf eine exakte Gesichtserkennung durch Gesichtsveränderung wie Mimik und Emotionen, einfallendes Licht oder störende Schatten, Abwendung der Blickrichtung von der Aufnahmekamera, oder Verschleierung beziehungsweise teilweise Verdeckung des Gesichtes verschlechtert werden.

Diese unvorteilhafte Umgebung entspricht jedoch dem häufig realen Bild, was besonders bei der Auswertung von Aufnahmen aus Überwachungskameras auffällt.

Die hier aufgezeichneten Gesichter entsprechen in der Regel keineswegs einem Idealfoto, da die Personen aus einem schrägen Winkel von oben herab aufgenommen, ihre Gesichter von Brillen und Hüten, Tüchern und Kapuzen verdeckt werden und Emotionen durch Lachen, Weinen und wütendes Stirnrunzeln wiederspiegeln.

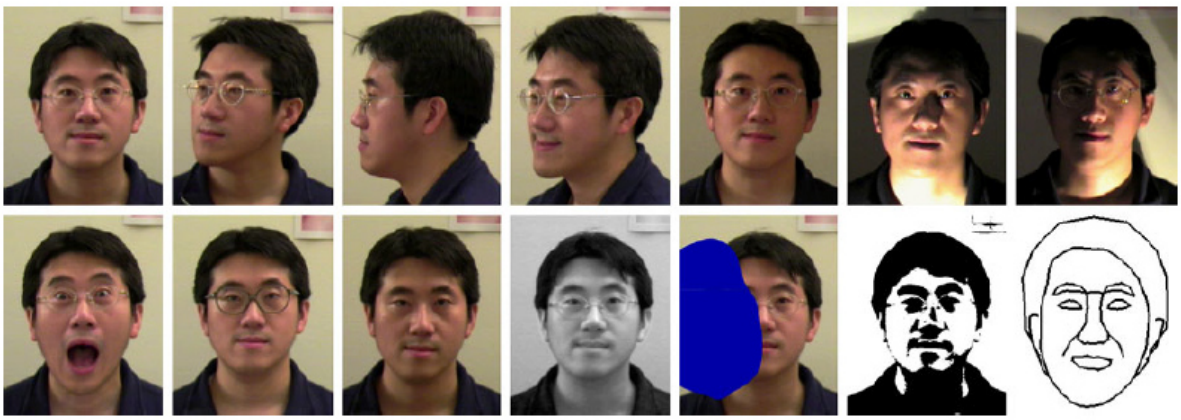


Abb. 1 Einfluss von Blickwinkel, Mimik, Beleuchtung, Teilverdeckung, Accessoires und Abstraktion auf die Varianz eines Gesichtes, aus Handbook of Face Recognition, [2], Eigentum von Rein-Lien Hsu[3]

Außerdem verschlechtern Schatten und blendendes Gegenlicht die Qualität der Aufnahmen. Einen umfangreichen dennoch nicht abschließenden Einblick in die Problematik der "unconstrained environments" verschafft Abb. 1.

Der standardmäßige Vorgang bei der Gesichtserkennung basiert auf vier Schritten: der Gesichtslokalisierung, der -normalisierung, der Feature-Extraktion und zu Letzt einer Feature-Übereinstimmungssuche [2 S. 4].

Bei der Gesichtslokalisierung oder auch Gesichtsfeststellung wird das vorliegende Bild oder Video zuerst dahingehend überprüft, ob überhaupt ein Gesicht enthalten ist. Die zugrundeliegende Mustererkennung sucht nach bestimmten Landmarken wie Augen, Nase, Mund und Haaransatz im Bildmaterial und findet bei modernen Implementierungen dementsprechend auch Gesichter, die zum Beispiel zur Seite gedreht oder teilweise verdeckt sind. Bei Videos besteht die Möglichkeit das Gesicht über mehrere Einzelbilder hinweg zu verfolgen und anschließend zusammenzusetzen.

Daraufhin folgt die Gesichtsnormalisierung, die weniger bei kooperativen Benutzerszenarios zum Einsatz kommt, dafür aber eine umso entscheidendere Rolle bei unkooperativen Benutzerszenarios innehält. In diesem Schritt werden die im Gesicht festgestellten Landmarken bei gleichbleibender Position so angeordnet, dass das Gesicht in "Normalposition", also gerade in die Kamera schauend, vorliegt. Mittels modernster Morphing-Methoden wird das Gesicht geometrisch so gedreht, als ob es in eine andere Richtung schauen würde. Dabei können fehlende Teile, wie zum Beispiel eine von der Kamera abgewandte Seite rekonstruiert und von Schatten verdunkelte Bereiche aufgehellt werden. Anschließend erfolgt auf Grundlage des normalisierten Gesichtes eine Feature-Extraktion, die dem Gesicht bedeutende, hervorstechende Merkmale entnimmt, bei denen nachweislich festgestellt wurde, dass sie zur eindeutigen Unterscheidung von Personen zielführend waren und daher umfassend eingesetzt werden, weil sie sehr divergent auftreten und robuste Trefferwahrscheinlichkeiten erzielen [2 S. 4].

Diese "Landmarken" werden im letzten Schritt, der Überprüfung nach Übereinstimmungen mit Features der Gesichter aus der Vergleichsdatenbank, genutzt. Je nach dem Vorhandensein im Datensatz, den Rahmenbedingungen, den Genauigkeitsregeln und den Erkennungsalgorithmen kann jetzt das erkannte und normalisierte Bild dem entsprechenden Bild aus der Datenbank zugeordnet werden oder den Bildern, die ihm am jeweils ähnlichsten sehen.

Die Genauigkeit und Fehlerraten aller Systeme hängen von den verwendeten Features ab, wie und ob das Gesicht erfolgreich normalisiert wurde und inwiefern die anfänglichen Mustererkennungen das Gesicht überhaupt festgestellt haben. Eine Optimierung der grundlegenden Algorithmik führt demnach zu einer höheren Wahrscheinlichkeit, dass Gesichter, deren Erkennung unter Einfluss von Licht, Gesichtsausdruck, Blickwinkel und Verdeckung erschwert wird, dennoch einander positiv zugeordnet werden können. Um dies zu ermöglichen muss die Erkennungssoftware gewisse Toleranzbereiche erschaffen, in denen sich ein Gesicht "bewegen" kann, die es aber dennoch von einem Anderen abgrenzen. Allerdings muss dennoch sichergestellt werden, dass ein System dahingehend exakt und präzise arbeitet, um Falsch-Positiv-Übereinstimmungen wie bei Zwillingen, vergleiche Abb. 2, zu vermeiden.

Abb. 2 Fehlerquelle für falsch-positive Übereinstimmungen - Zwillinge, heruntergeladen von: http://images.medicaldaily.com/sites/medicaldaily.com/files/2016/05/23/twins.jpg

2.2.1 Grundlegende Methoden zur Gesichtserkennung

Das wohl naheliegenste Verfahren der Gesichtserkennung ist der direkte Vergleich zweier Bilder über eine Ähnlichkeitsfunktion.

Ein in der multivariaten Statistik verwendetes mathematisches Verfahren, das auch bei der Gesichtsrekonstruktion zum Einsatz kommt, ist die Hauptkomponentenanalyse, die auch als Principal Component Analysis (PCA) bekannt wurde.

Inhalt dieser ist es, eine große Anzahl von Datenvariablen in einer n-dimensionalen Punktwolke durch wenige Hauptkomponenten zu ersetzen, und damit umfangreiche Datensätze zu veranschaulichen. Diese Principal Components sind möglichst aussagekräftige Linearkombinationen, die an den statistischen Datensatz so angenähert wurden, dass allenfalls ein geringfügiger Informationsverlust auftritt. Zum Vereinfachen der Daten wird eine Anzahl an n Geraden durch den Mittelpunkt der Daten gesucht, deren durchschnittlicher Fehler am geringsten ist. Außerdem müssen die Geraden senkrecht zueinander stehen. Mathematisch korrekt sind diese Geraden eigentlich Vektoren, da sie neben einer Richtung auch einen genauen Wert besitzen.

Im weiteren Verlauf wird die Varianz der einzelnen Datensätze entlang der Geraden aufsummiert, wobei auffällt, dass die Varianz von Gerade zu Gerade sukzessiv geringer wird.

Jede Gerade verfügt über immer weniger Varianz als ihre Vorgänger, was einen bestimmten Anteil an Geraden im hinteren Segment, deren Anzahl je nach Versuchsaufbau unterschiedlich ausfällt, unerheblich oder gar irrelevant macht [2 S. 5, 20-23] .

Für die Gesichtsrekonstruktion bei Foto- und Videoaufnahmen wird vom sogenannten "face subspace" ausgegangen, das heißt, innerhalb des aus tausenden von Pixeln bestehenden Bildes (auch "image space" genannt) existiert nur ein Teil davon, der das eigentliche Gesicht ausmacht [2 S. 5, 20] [3, 4, 5, 8, 11, 15, 16]. Alle Gesichter besitzen eindeutige Merkmale, die zwar je völlig unterschiedlich ausgeprägt sind, aber in ihren Grundkomponenten stets vorkommen. Diese Merkmale entsprechen den Vektoren der allgemeinen PCA.

Bei einer Datenbank von Vergleichsbildern wird ein Durchschnittsgesicht mittels aller normalisierten Gesichter ermittelt, das daraufhin von jedem einzelnen Bild subtrahiert wird, um den Datensatz zu normieren. Definiert durch den Wert der Eigenvektoren entstehen nun neue Aufnahmen, die die Differenzen zum Durchschnittsgesicht darstellen. Sie werden als "eigenfaces" bezeichnet und sind in Abb. 3 zu erkennen. Je mehr Vektoren hinzugenommen werden, desto einfacher ist es, das Gesicht von anderen zu unterscheiden. In der Regel reichen die ersten zehn bis dreizehn Vektoren, da sie die größte Varianz besitzen und sich daher am besten eignen, ein Gesicht eindeutig zu identifizieren [2 S. 5-6] [4].

Wird zum Vergleich mit dem Datensatz ein weiteres Gesicht hinzugefügt, berechnet der Algorithmus die Euklidische Distanz (kürzeste Distanz durch den Raum) zwischen den Koeffizienten-Vektoren des Bildes und denen der Hauptkomponentenanalyse der eigenfaces. Das Ergebnis kann folgendermaßen klassifiziert werden: A) zu einem Gesicht gehörig (Übereinstimmung), B) nicht zu einem Gesicht gehörig (neues, nicht archiviertes Bild), oder C) kein Gesicht enthaltend. Hierbei wird deutlich, dass sich diese Methode nicht nur zur Unterscheidung von Gesichtern sondern sogar auch zur Entdeckung ebendieser eignet.

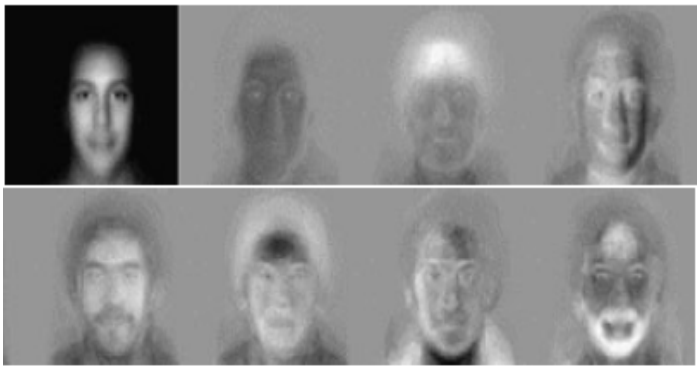


Abb. 3 Links oben: Durchschnittsgesicht, rechts oben und untere Zeile: sieben Beispiele für "eigenfaces", , aus Handbook of Face Recognition[2], Eigentum von Turk und Pentland[4]

Es existieren verschiedene Abstraktionen der PCA und Herangehensweisen, die auf ihr beruhen. Dazu zählen untere anderem die ICA (Vergleich mit PCA in Abb.4), Independent Component Analysis, bei der die Vektoren nicht zwangsläufig senkrecht aufeinander stehen müssen und die Reihenfolge der ausgewählten Vektoren abweichen kann [2 S. 29-31], und daher den Datensatz möglicherweise besser auffangen können, sowie die KPCA, die den Standard-PCA-Algorithmus um das Kernel-Element erweitert [2 S. 10, 35-36].

Bei der Kernel Principal Component Analysis handelt es sich um eine Methode mit der lineare Klassifikationen auf nicht-lineare Daten angewendet werden können, da sie in einen Raum transformiert werden, der aus einer höheren Anzahl Dimensionen besteht. Dort werden die Daten implizit verarbeitet, das heißt es werden nicht jeweils sämtliche Koordinaten der Ausgangsdaten in die Berechnung der Vektoren übernommen, sondern nur jeweils Skalarprodukte zwischen den einzelnen Datenpunkten berechnet. Der KPCA gelingt es Falsch-Positiv-Treffer gering zu halten und kommt ohne Vorkenntnisse über die Dimensionsanzahl aus.

Außerdem ist es möglich mehr Vektoren zu erstellen als die Dimensionalität der Ausgangsdaten vorgibt. Allerdings stellt die optimale Auswahl des Kernels ein ebenso komplexes und datensatzabhängiges Problem wie die Anzahl der verwendeten Vektoren der PCA dar [2 S 35-36].

Des Weiteren existiert die bereits seit mehreren Jahren in der Gesichtserkennung verwendete LDA, Linear Discriminant Analysis, eine Erweiterung der Standard-PCA um eine möglichst optimale lineare Funktion, deren Aufgabe es ist, die Ausgangsdaten in den Klassifikationsraum zu übertragen. Dieser besteht aus den einzelnen Klassen der Daten und modelliert explizit die Unterschiede zwischen den Klassen. Dadurch hilft LDA der Gesichtserkennung im Besonderen dadurch, dass es Variationen in Datensätzen den einzelnen Klassen (und damit den Identitäten der Personen) zuordnet.

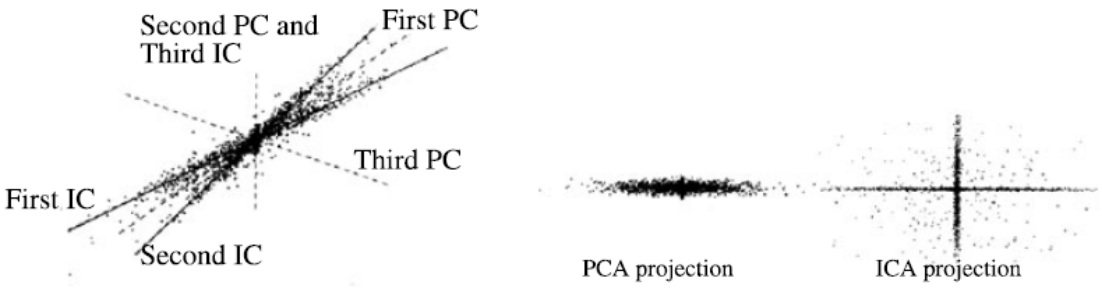


Abb. 4 links: Punktwolke mit Geraden aus PCA und ICA (in Reihenfolge), rechts: PCA und ICA Projektionen, aus Handbook of Face Recognition[2]

Bei einer Operation nur mittels PCA kann es passieren, dass Variationen der gleichen Art untereinander eher zugeordnet werden als der zugehörigen Identität mit anderer Variation. Das bedeutet, dass zwei Bilder mit lachenden Gesichtern einander zugeordnet werden, weil die gleiche Art der Variation (hier: lachende Mimik) vorliegt, aber nicht den jeweils zugehörigen Personen [2 S. 7-11] [6].

Dies versucht die LDA voneinander getrennte Klassen zu umgehen und nutzt dafür die Fisher'sche Diskriminanzfunktion (optimale Trennung der Klassen durch vorherige Dimensionsreduzierung), weswegen die klassifizierten Bilder in der Literatur gerne als "Fisherfaces" (in Analogie zu "eigenfaces") tituliert werden. Bei vielen Versuchen mit LDA zur Gesichtserkennung traten jedoch hohe Fehlerraten auf, wenn das verwendete Gesicht auf dem untersuchten Bild nicht in der vorherigen Trainingsdatenbank vorhanden

war, sich die Bilder durch verschiedene Hintergründe unterschieden oder die Beispiele der Trainingsklassen extrem stark voneinander abwichen [6].

Eine der modernsten Methoden, die in Studien der letzten Jahre deutlich bessere Ergebnisse erreichte als die PCA und ihre Ableger [2 S. 39-42] [5], ist die Bayes'sche Methode und ihre Berechnung der bedingten Wahrscheinlichkeit [2 S. 27-29] [5] [7 S. 37-39, 100-101].

Im Gegensatz zu den zuvor beschriebenen Methoden, basiert diese nicht auf der Euklidischen Distanz und der sogenannten "nearest-neighbour based recognition"[3], also der Erkennung durch die nahegelegensten Nachbardaten. Dieses relativ einfache metrische Konzept leidet unter dem Makel, dass es die Art der vorliegenden Gesichtsvariation nicht unterscheiden kann [5] [7 S. 100].

Der Bayes'sche Ansatz kann jedoch zwischen Intrapersonellen Abweichungen (Unterschiede zwischen Bildern mit demselben Gesicht, abweichend in Beleuchtung und Expression) und Extrapersonellen Abweichungen (Unterschiede zwischen Bildern verschiedener Personen) unterschieden, die über Wahrscheinlichkeitsberechnung mittels des Satz von Bayes errechnet werden. Es wird davon ausgegangen, dass die Ähnlichkeit zweier Bilder bedingt abhängig von der Intra- und Extrapersonellen Variation ist. Ziel dieser prohabilistischen Vorgehensweise ist eine eindeutige Gesichtserkennung und ein Abgleich der Datenbankeinträge mit höherer Übereinstimmungsgenauigkeiten, trotz der Verwendung von nicht-idealen Bildern, die durch Mimik, Lichteinfluss und Blickwinkel verändert wurden. Dabei ist das Verfahren so präzise, dass es im "unconstrained environment" voll-operativ ist, höhere Ergebnisse erzielt als PCA, IDA und KPC und den Datenbankspeicherbedarf senkt, da es pro Person nur genau ein Bild benötigt (und nicht mehrere Bilder unter verschiedenen Bedingungen) [5].

Neben den hier vorgestellten state-of-the-art Methoden gibt es viele weitere Abstraktionen, Verknüpfungen und völlig unabhängige Vorgehensweisen zur Gesichtserkennung und -unterscheidung. Es sind folglich unzählige Versuche und Experimente mit unterschiedlichsten Datenbanken durchgeführt worden, die die jeweiligen Vor- und Nachteile der einzelnen Methoden und ihre Erfolgsabhängigkeit vom verwendeten Datensatz protokollieren.

---

[3] Moghaddam und Pentland, Bayesian Face Recognition, S. 2 [5]

Alle beschriebenen Methoden werden auch in angepasster, erweiterter und kombinierter Form bei der Alters- oder Geschlechtsbestimmung genutzt, die auf der ursprünglichen, reinen Gesichtserkennung aufbauen, da die Methoden einander überschneiden und moderne Systeme sowohl Gesichtserkennung als auch Alters- und Geschlechtsbestimmung beherrschen. Die nachfolgend beschriebenen Methoden gelten als geeignet und werden wie beschrieben genutzt. Nichtdestotrotz ist diese kurze Zusammenfassung nicht abschließend, und viele Methoden finden auch außerhalb des hier erklärten Bereiches ihre Anwendung.

2.2.2 Altersbestimmung

Die präzise Altersbestimmung oder zumindest eine möglichst genaue Schätzung des Alters einer Person, nur anhand ihres Gesichtes auf einer Videoaufnahme, stellt innerhalb der Gesichtserkennung nochmals einen komplexen Unterbereich für sich dar. Das Alter einer Person unterliegt unkontrollierbaren Variablen, die nur schwer nachvollziehbar sind und die selbst das menschliche Auge nur minderpräzise abschätzen kann. Der Altersprozess eines jeden Menschen verläuft geringfügig anders, individuelle sowie geschlechterspezifische Wachstumsprozesse setzen zu unterschiedlichen Zeiten ein. Dazu kommen Phasenschübe wie in der Pubertät und externe Faktoren, die auf das augenscheinliche Alter einwirken, wie arbeitsbedingter Stress, die Einnahme von Droge oder langzeitiges Rauchen.

Ebenso werden entscheidende Details, die zur Bestimmung des Alters benötigt werden, von uns mehr oder weniger bewusst versteckt, beispielsweise durch Bärte, Brillen und vor allem Make-up und Hygieneprodukte, deren primäre Aufgabe es ist Zeichen des voranschreitenden Alters, wie Falten, zu überdecken [8, 9, 17].

Eine Vorgehensweise, die sich bei der automatischen Altersschätzung bewährt hat, ist die Extraktion von Gesichtslandmarken oder auch "feature extraction" [8, 11]. Vorrausetzung dieser Methode ist, dass diese Landmarken (engl. "features") diskriminativ innerhalb der Klassen sind, innerhalb der Klasse jedoch relativ robust bleiben, und eine geringe Dimensionalität aufweisen [8].

Es müssen einige aussagekräftige Parameter gefunden werden, mit deren Hilfe Gesichter eindeutig beschrieben werden können.

Durch optimierte quadratische Funktionen werden Beziehungen zwischen den Gesichtsparametern und dem Alter der Person erstellt. Außerdem wird ein Globaler Altersschätzungsklassifikator ermittelt, dessen Funktion eine vorerst recht grobe Unterteilung in größere Altersgruppen mit einem mehrjährigen Betrag ist; häufig werden hierfür Intervalle zwischen fünf und zehn Jahren genutzt.

Innerhalb dieser Gruppen führt man anschließend die weitaus präzisere, lokale Altersbestimmung durch.[10, 11, 12].

Die beiden bekanntesten Vertreter dieser Methodik sind die aufeinander basierenden Modellarten Active Appearance Model (AAM) [8, 12, 13, 9] und Active Shape Model (ASM) [8, 14, 9]. Der Vorgänger ASM hat mittels der untereinander verbundenen Punkte einer Punktwolke den Umriss eines Objektes beschrieben. Das Model arbeitet mit einer einheitlichen Form der Landmarken, anhand dieser es Abweichungen als Gesichtsvariation erkennt und alle Bilder in eine normierte Standardform konvertiert. Die Erweiterung AAM erkennt zusätzlich auch Grauwertverteilungen auf dem Gesichtsbereich. Beide Formen sind "Active", da die Parameter automatisch angepasst werden, sobald ein oder mehrere Bilder, die dem Datensatz hinzugefügt werden, eine bessere Anpassung des Modells bedeuten. Sie sind also in der Lage durch die Daten zu lernen [12, 13]. Jedoch kann der hochkomplexe natürliche Vorgang des Alterns nicht mit einer einfachen Quadratfunktion beschrieben werden, die noch dazu das Alter aller Gesichter fälschlicherweise am Altersprozess einiger weniger Personen des Trainingsdatensatzes berechnet [17].

Deshalb findet neben der reinen Landmarkenextraktion auch das maschinelle Lernen durch vorab Klassifikation und Regression ihre Anwendung bei der Altersbestimmung, da es sich als adaptiver herausgestellt hat und die umständliche und ungenaue Unterteilung in Altersintervalle umgeht [15].

Es existieren verschiedene Familien des regressionsbasierten, maschinellen Lernens: Support Vector Regressionen, (mehrschichtige) Neuronale Netzwerke [11] und Projektionsanalysen, wie Canonical Correlation Analysis und Partial Least Squares [8, 15].

CCA sucht nach Basisvektoren, beruhend auf den Abhängigkeiten der Inputvariablensätzen, sodass die Projektionen dieser Vektoren maximal miteinander korrelieren, siehe Abb. 5.

Des Weiteren kann CCA genutzt werden um die Dimensionsanzahl zu reduzieren, durch die Berechnung der Variablen, die die höchste Korrelation besitzen [18, 19]. Auch hier existiert die Untergattung KCCA, die Kernelbasierte Canonical Correlation Analysis. Ihr Zweck ist die Anpassung der hypothetischen Vektoren in einem höher dimensionierten Raum, durch den die Flexibilität an die Varianz der Daten erhöht wird [19], was vor allem dann von großer Bedeutung ist, wenn die Ausgangsdaten nicht linear sind [18]. KCCA senkt ebenso den Berechnungsaufwand und erlaubt die Übertragung eines einzelnen Validierungsschemas auf alle anderen Daten innerhalb einer beliebig großen Datenbank, ohne dabei an eine Größengrenze gebunden zu sein [15].

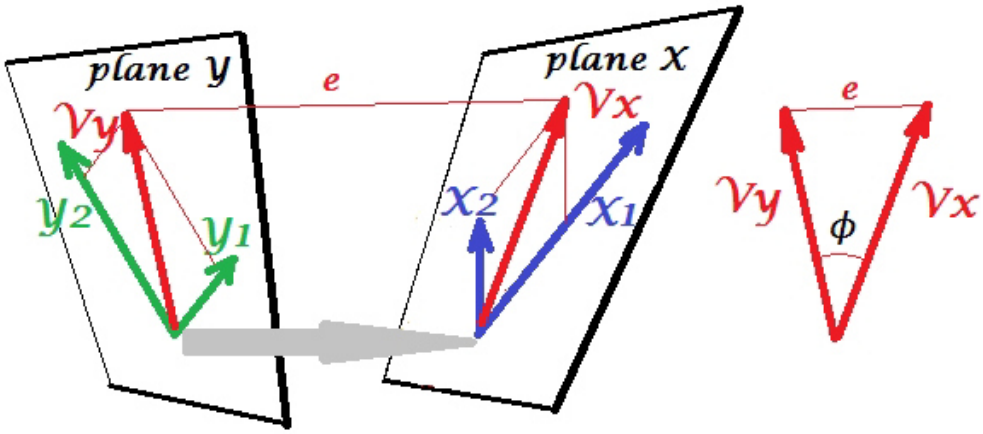


Abb. 5 Visualisierung CCA - Bildung der neuen Basisvektoren (rot) aus zwei Variablensätzen (grün und blau) , heruntergeladen von: https://i.stack.imgur.com/D9dbY.jpg

Partial Least Squares hingegen arbeitet mit latenten Variablen um einen neuen Teilbereich, "subspace" zu definieren, in dem die vorausgesagten Variablen mit den vorgegebenen des Ausgangsdatensatz maximale Korrelation besitzen [8, 15, 16]. PLS erweist sich als besonders effizient wenn die Input-Daten eine hohe Multikollinearität aufweisen, das bedeutet die Variablen korrelieren untereinander sehr stark, sodass die Schätzung der Regression und ihrer -koeffizienten uneindeutig und ungenau wird.

Eine sehr moderne Praxis der Altersbestimmung ist die Operation mit einem visuellen HOG-Deskriptor in Kombination mit einer Projektionsausrichtung von genau fünf Landmarken (beide Mundwinkel, beide Pupillen und Nasenspitze) [15]. Bei HOGs, Histogram of Oriented Gradients, handelt es sich um robuste Erkennungsdeskriptoren, die dank ihrer schnellen Verarbeitungszeit, Kompaktheit und geringen Varianz durch Beleuchtungsunterschiede und

Falschausrichtungen geeignet für die Gesichtserkennung und Altersbestimmung sind [15]. Für das Histogramm wird das Basisbild in 50x50 Pixel große, gleichförmige, quadratische Teilbilder separiert. Innerhalb eines jeden lokalen Bereiches wird für die vorhandenen Pixel je ein Richtungsgradient kompiliert, siehe Abb. 6. Der finale Deskriptor ist letztendlich die Verkettung der einzelnen Gradienten-Histogramme [15].

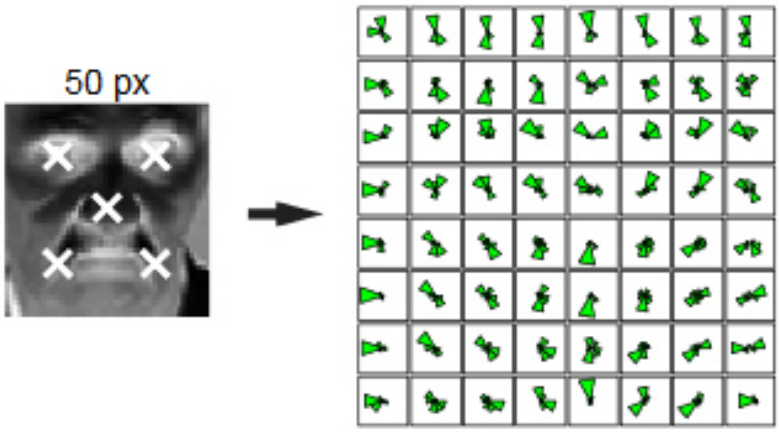


Abb. 6 50x50 Pixel großes Teilbild, daraus resultierendes Histogramm aus Richtungsgradienten, aus A Comparative Evaluation of Regression Learning Algorithms for Facial Age Estimation, [15]

Die nach Dennis Gabor benannten "Gabor Filter" bilden eine weitere, umfangreiche Transformationsstufe, die nach bestimmten Mustern sucht und die sowohl eigenständig aber auch vorangestellt an zuvor beschriebenen Methoden angewendet werden kann. Gabor Filter (GF) werden in der Bilderverarbeitung zur Kantendetektion genutzt, mit deren Hilfe unterschiedliche "subspaces" voneinander getrennt sind und deshalb auch automatisch unterschieden werden können. Dabei sind GF an das natürliche visuelle System des Menschen angelegt und funktionieren entsprechend ähnlich.

Prinzipiell wird die Bildung eines GF definiert durch eine sinusförmige Funktion oder eine ebene Welle (2D), die mit einer Gauß'schen Funktion multipliziert wurde. An ausgewählten Landmarkenstandorten im Gesicht werden diese sogenannten "Wavelets" angewendet, um einen Vektor zu errechnen, der das Gesicht durch eine große Anzahl diskrimininanter Informationen charakterisiert [9, 20, 21]. Die verschiedenen Wellen treffen an unterschiedlichen Punkten aufeinander, an denen sich bereichstrennende Kanten bilden.

Um einzelne Kanten, die in verschiedene Richtungen verlaufen, zu erkennen, können mehrere Gabor Filter in einer Rotation miteinander verknüpft werden.

Das aus der rotierten Verknüpfung entstandene Bild nennt man "Gabor Space". Dieser beinhaltet sämtliche Kanten, die jeweils einzelne Teilbereiche voneinander trennen, wie in Abb. 7 erkennbar ist. Man spricht hierbei von einer "image decomposition" oder Bildzerlegung [20].

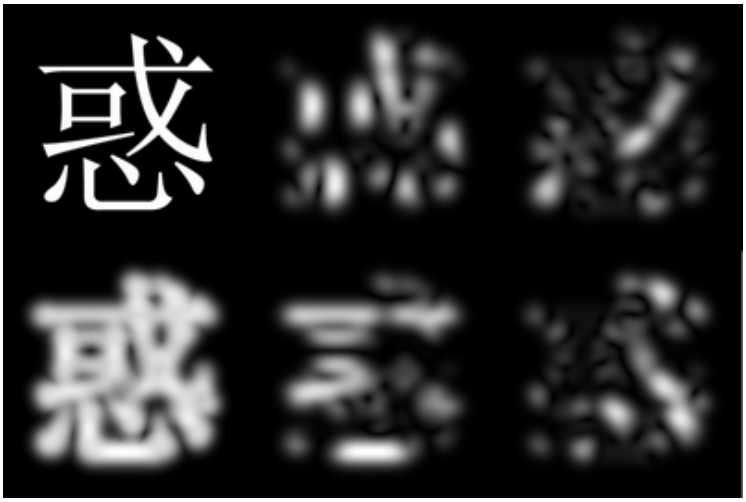


Abb. 7 Demonstration einer Gabor Filter Kombination durch Rotation bei der Erkennung eines chinesischen Schriftzeichens. links oben: Ausgangsbild, rechts: vier Erkennungsmuster aus verschiedenen Winkeln (0°, 45°, 90°, 135°), links unten: Zusammengesetzes Endbild, heruntergeladen von: http://cdn-ak.f.st-hatena.com/images/fotolife/Z/Zellij/20131003/20131003181044.png

Alle diese Methoden der Altersbestimmung haben eines gemeinsam: Sie klassifizieren das Alter der Personen anhand ihrer mathematischen Berechnung (Vektoren, Gradienten, Regressionen) und erstellen einen grundlegenden Ablauf des Alterungsprozesses. Bildlich gesprochen gehen die Gesichtsbilder eines Menschen mit zunehmendem Alter sukzessiv von Status zu Status über. Über alle Personen hinweg wird eine Funktion aufgestellt mit der man Alter berechnen kann. Wird ein neues Vergleichsbild dem Datensatz hinzugefügt, dann wird überprüft welchem Status es entspricht.

Eine letzte noch zu erwähnende Entwicklung in der Altersbestimmung befasst sich mit dem sogenannten Altersmuster. Der Inhalt dieser Methode ist der Vergleich von Bildern ein- und derselben Person in verschiedenen Altern miteinander, siehe Abb. 8. Es wird verstärkt Bezug zum individuellen Altersprozess gemacht, anstatt wie die meisten anderen Vorgehensweisen mittels verschiedener Personen verschiedenen Alters einen Klassifikator für Alter zu ermitteln [17].

Die Methode des Altersmuster interessiert sich für den fortwährenden Altersprozess an altersbedingt veränderlichen Landmarken und generellen Merkmalen des zunehmenden Alters wie Falten und fehlende Hautstraffung, das heißt für die individuelle Veränderung jedes einzelnen Gesichtes [17].

Die Methode hat sich als sehr präzise herausgestellt, erfordert jedoch umfassende Datensätze mit Gesichtsbildern jeder Person aus möglichst vielen verschiedenen Lebenszeitpunkten. Letztendlich läuft der Alterungsprozess einer jeden Person differenziert ab und so können individuelle Vorgänge und damit eine höhere Varianz in das Modell einbezogen werden [17].

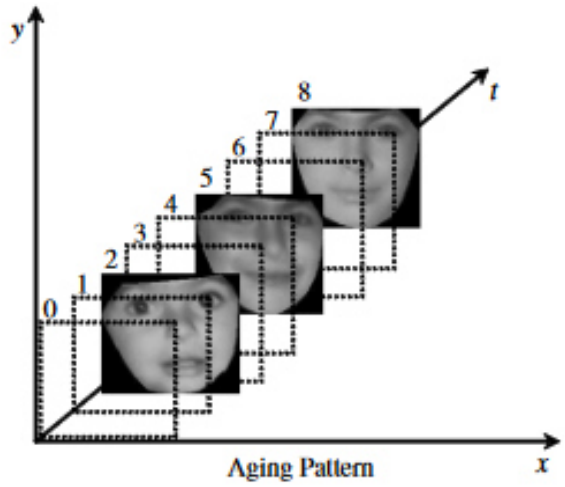


Abb. 8 Altersmuster aus Bildern einer Person zu verschiedenen Zeitpunkten, aus Learning from Facial Aging Patterns for Automatic Face Estimation [17]

2.2.3 Geschlechtsbestimmung

Auch die Bestimmung des Geschlechtes nur anhand des Gesichtes ist eine beträchtliche Herausforderung für die automatische Erkennungssoftware, noch dazu weil in modernen Zeiten selbst dem menschlichen Auge die eindeutige Unterscheidung von Mann und Frau durch Make-Up, untypische Haarschnitt-Trends (lang bei den Männern, kurz bei den Frauen) und der zunehmenden gesellschaftlichen Akzeptanz von Transsexualismus und Geschlechtsoperationen erschwert wird [23, 26].

Eine der bekanntesten und am umfangreichsten getesteten Varianten zur Geschlechtserkennung und -unterscheidung ist die Methodik der Local Binary Pattern (LBP), der lokalen Binärmuster. Diese umfasst die Unterteilung des Gesichtes in gleichgroße Quadrate, standardmäßig 16*16, 30*30 oder 50*50 Pixel groß. Jedes Pixel wird mit seinen acht Nachbarn (soweit vorhanden), also den Pixeln rechts, links, oben, unten, rechts oben, links oben, rechts unten und links unten, verglichen.

Dieser Vorgang wird im Normalfall im Uhrzeigersinn durchgeführt und bei jedem Vergleich wird eine Binärzahl geschrieben: eine Null, wenn der Wert des zentralen Pixels größer ist als der, des zu vergleichenden Nachbars, andernfalls eine Eins. Insgesamt entsteht so eine achtstellige Binärzahl. Nun kann ein Histogramm erstellt werden, das aus den Anzahlen der jeweiligen Binärwerte der einzelnen Pixel besteht. Schlussendlich wird das Histogramm normalisiert und verkettet, sodass ein Gesamtvektor entsteht [22, 23].

Die Vorgehensweise des LBP ist als direkter Vorgänger zum schon erwähnten HOG zu sehen, unterscheidet sich in der Verwendung der Binärzahlen anstatt der Richtungsgradienten und ist in der Geschlechtsbestimmung weiter verbreitet [22, 23, 26].

Beide wurden jedoch genutzt um die geschlechterabhängige Gesichtsform zu untersuchen, denn es hat sich herausgestellt, dass statistisch gesehen Männer im Allgemeinen eher elliptische Köpfe haben, während die Gesichter der Frauen meist eine eher rundliche Form besitzen [26]. Außerdem werden durch LBP und HOG deutliche Unterschiede zwischen männlicher und weiblicher Hautbeschaffenheit und große Differenzen im Augen- und Augenbrauenbereich ersichtlich [26].

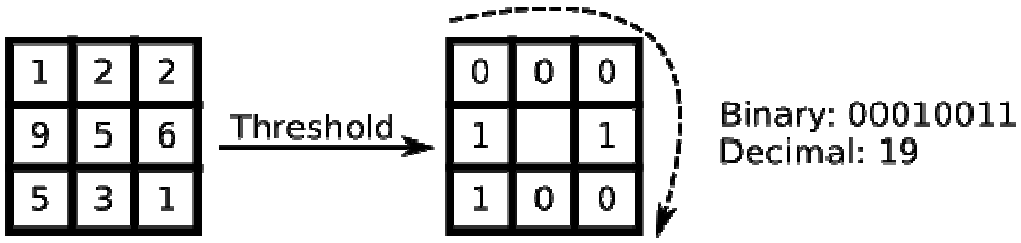


Abb. 9 Visualisierung LBP: Beispielpixelwerte links, Binäre Umwandlung, heruntergeladen von: http://bytefish.de/static/images/blog/local_binary_patterns/lbp.png

Eine Erweiterung stellt die Zwei-Klassen-Unterteilung nach Intra- und Extrapersonellen Gesichtsfeatures dar; erstere werden beim Vergleich von Bildern derselben Person genutzt, letztere wenn mit anderen Personen verglichen wird. Unterscheidungen werden durch das Chi-Quadrat-Entfernungsmaß berechnet [24].

Um zuverlässige Ergebnisse zu erreichen werden die erhaltenen Daten mittels Boosting kompiliert. Dabei werden mehrere schwächere Klassifikatoren kombiniert um präzise, starke Klassifikatoren zu erhalten, die eine erheblich höhere Genauigkeit und einen verringerte Fehlerquote aufweisen [2 S 2,11] [22, 24]. Der bekannteste Boostmechanismus ist der von Freund und Schapire [25] beschriebene AdaBoost, der, ganz im Gegensatz zu anderen Boostingverfahren, keinerlei Vorkenntnisse über die schwachen Klassifikatoren benötigt, sondern sich diesen anpasst und einen neuen Gewichteten kreiert [25].

Es existieren weitere Verfahren der Geschlechtsbestimmung, vor allem durch maschinelles Lernen mittels Support Vektor Regressionen und Neuronale Netzwerke. Des Weiteren können diverse Methoden angewendet werden um kombinierte Versuche, die sowohl Gesichtsdetektion und -erkennung als auch Alters- und Geschlechtsbestimmung beherrschen, durchzuführen [16, 27, 28].

2.3 SHORE

Die wohl bekannteste und auch die am häufigsten eingesetzte Standardbibliothek für Gesichtserkennungsalgorithmik ist OpenCV, die Open Source Computer Vision Library. Als eine kostenlos erhältliche Infrastruktur für Anwendungen im Bereich der automatischen kognitiven Services und des maschinellen Lernens, dient sie als Grundlage vieler kommerzieller Produkte. Sie besteht aus mehr als 2500 optimierten Algorithmen, die Gesichter, Gegenstände und Schriftzüge oder Logos erkennen, identifizieren, vergleichen, deren Bewegungen verfolgen und in 3D-Formate überführen können.

Die Bibliothek wurde über 14 Millionen mal heruntergeladen und wird von großen Unternehmen wie Google, Microsoft, Sony und Intel genutzt[4].

Viele der in dieser Arbeit erwähnten Programme, besonders zur Gesichtserkennung, beruhen auf OpenCV und implementieren die bereitgestellten Algorithmen. Auch das hier verwendete SHORE, basiert auf OpenCVs Erkennungsfunktionen.

OpenCV stellt ein komplexes Framework dar, dessen Funktionen umfassend sind. Es wäre möglich basierend auf den Algorithmen, eine eigene Software zur Gesichtserkennung mitsamt Alters- und Geschlechtsbestimmung zu entwickeln. Ziel dieser Arbeit ist es jedoch bereits vorhandene, teils von populären Firmengiganten und renommierten Instituten vertriebene Lösungen zu testen.

Das Fraunhofer Institut für Integrierte Schaltungen (IIS) arbeitet schon seit 2005 an einer Software zur automatischen Gesichtserkennung, die den Titel SHORE, abgekürzt für Sophisticated High-speed Object Recognition Engine, trägt. Anfangs stellte das Tool nur die Möglichkeit zur Gesichtsdetektion zu Verfügung, also die reine Untersuchung eines Bildes nach Gesichtern. Wurde ein Gesicht entdeckt, und demnach die Landmarken erkannt, wurde es im Bild markiert. Über die Jahre kam mit verschiedenen System- und Erkennungsoptimierungen auch die Fähigkeit zur zeitglichen Feststellung mehrerer Gesichter hinzu. Außerdem war es später möglich neben Bilderformaten auch Videodateien und den Livefeed der Kamera zu analysieren.

Erst im Jahr 2010 wurde aus der Gesichtsdetektion eine Gesichtserkennung, da nun das Gesicht einer Person zugeordnet werden konnte, sowie Alters und Geschlechtsbestimmung Einzug in die Softwarelösung erhielten. Zu guter Letzt wurde die Emotionsanalyse hinzugefügt, die anhand von Augen-, Stirn- und Mundneigungen die Stimmung der Person angeben konnte. Seither wird das Programm mit modernen Algorithmen verbessert und stabilisiert.

[4] OpenCV team, "About - OpenCV library", http://opencv.org/about.html, angesehen am 16. Mai 2017

Es wurden Tests mit Bildern aus populären Portraitdatenbanken, wie CMU+MIT, BioID und FG-NET durchgeführt. Laut eigenen Angaben erzielt SHORE dabei eine Performance von 91,5% bis 94,3% sowie ein durchschnittliche Altersschätzungsdifferenz von 6,85 Jahren. Für eine vollständige Analyse inklusive Emotions-, Alters- und Geschlechtsbestimmung benötigt SHORE 22 Millisekunden für ein Bild und schafft 45,5 Bilder pro Sekunde[5].

Internationale Aufmerksamkeit erlangte das Programm des IIS im Jahr 2014 als es in Kooperation mit der Datenbrille Google Glass vorgestellt wurde und die in die Brille integrierte Kamera nutzte, um einen Livefeed von Personen, die dem Träger gegenüberstanden, hinsichtlich Alter, Geschlecht und Stimmungslage zu analysieren.

Für die im Rahmen dieser Arbeit angestrebten Versuche hat sich SHORE allen anderen, vergleichbaren Softwarelösungen überlegen gezeigt. Einerseits ist die Analyse von Videomaterial, das nicht vor der Untersuchung segmentiert werden muss, ein zeitsparender und ressourcenschonender Vorgang. Anderseits arbeitet SHORE lokal und muss die zu untersuchenden Daten nicht in eine Cloud hochladen. Des Weiteren gibt die Altersanalyse einzelne Jahre an, nicht nur grobe Altersgruppen.

Die Unterschiede zu weiteren Softwarelösungen werden in Kapitel 5.1 näher erläutert.

[5] Fraunhofer IIS, "SHORE®", https://www.iis.fraunhofer.de/en/ff/sse/ils/tech/shore-facedetection.html, angesehen am 17. Mai 2017

3 Methoden

3.1 Auswahl der Testdateien

Um verfahrensähnliche Umstände zu suggerieren und entsprechend relevante Altersabschnitte (wie die Grenze zwischen Kindern (<14 Jahre) und Jugendlichen und die zwischen Jugendlichen und Erwachsenen (>=18 Jahre)) untersuchen zu können, ohne jedoch auf inkriminiertes Material zurückzugreifen, werden Stellen aus allgemein bekannten Filmen und Serien verwendet, die mehrere Szenen mit Schauspielern im Kindes- oder Jugendalter aufweisen. Es werden ebenfalls einige Szenen hinzugefügt mit Schauspielern, die knapp über dem relevanten Alter liegen, also junge Erwachsene zwischen 18 und 19 Jahren.

Alle Szenen werden gezielt so ausgewählt, dass die Schauspieler nicht nur als Statisten im Hintergrund stehen oder nur flüchtig im Bild erscheinen, sondern den aktiven Handlungsträger darstellen und bis zu mehrere Minuten "Screentime" einnehmen. Dabei sollte es sich möglichst um zeitnahes Material handeln, um durch technologiebedingte Qualitätsverbesserungen die Gesichtserkennungsprozesse zu optimieren.

Insgesamt handelt es sich um 54 Dateien, aus 18 Filmen und sieben Serien, mit 15 Schauspielern und zehn Schauspielerinnen (zwei treten in je zwei unterschiedlichen Filmen mit unterschiedlichem Alter auf und werden daher fortführend doppelt aufgeführt) im Alter von fünf bis 19 Jahren. Jede Person besitzt zwei Clips und sie werden jeweils unabhängig voneinander bewertet. Die Gesamtlänge der Videos beläuft sich auf 18 Minuten und 41 Sekunden, die durchschnittliche Länge der Clips beträgt circa 20 Sekunden. Diese Dauer wurde gewählt, um den Analyseprogramm genügend Material zur Gesichtserkennung sowie Geschlechts- und Altersbestimmung zu verschaffen. Je nach Material sind jedoch selbst Sekundenbruchteilen genügend, da sie aus aussagekräftigen Einzelbildern bestehen.

Folgend in Tabelle 1 aufgelistet, befinden sich eine Übersicht über die zusammengeschnittenen Videos mitsamt Informationen zu Alter und Geschlecht der jeweiligen Schauspieler.

Datei-name .avi	**Länge**	**Filmtitel**	**Schauspieler (-innen)**	**Geschlecht**	**Geboren am**	**Alter zum Drehzeit-punkt**
Dat1	00:33	Tatsächlich Liebe	Thomas Brodie-Sangster	Männlich	16.05.1990	13
Dat2	00:35					
Dat3	00:30	Keinohrhase	Emma Schweiger	Weiblich	26.10.2002	5
Dat4	00:11					
Dat5	00:19	Honig im Kopf	Emma Schweiger	Weiblich	26.10.2002	12
Dat6	00:09					
Dat7	00:16	Interstellar	Mackenzie Foy	Weiblich	01.11.2000	14
Dat8	00:16					
Dat9	00:24	Das Dschungelbuch	Neel Sethi	Männlich	22.12.2003	13
Dat10	00:23					
Dat11	00:20	Kevin Allein Zu Haus	Macaulay Culkin	Männlich	26.08.1980	10
Dat12	00:16					
Dat13	00:20	Karate Kid	Jaden Smith	Männlich	08.07.1998	12
Dat14	00:30					
Dat15	00:12	Game of Thrones 3.02	Maisie Williams	Weiblich	15.04.1997	16
Dat16	00:20					
Dat17	00:13	Hilfe, es weihnachtet sehr	Johnny Galecki	Männlich	30.04.1975	14
Dat18	00:07					
Dat19	00:14	The Walking Dead 7.01	Chandler Riggs	Männlich	27.06.1999	17
Dat20	00:24	The Walking Dead 7.04				
Dat21	00:13	The Strain 2.01	Max Charles	Männlich	18.08.2003	12
Dat22	00:10					
Dat23	00:18	Harry Potter und der Stein der Weisen	Emma Watson	Weiblich	15.04.1990	11
Dat24	00:33					
Dat25	00:14	Jurassic Park	Joseph Mazzello	Männlich	21.09.1983	10
Dat26	00:14					
Dat27	00:27	Star Wars Episode 1	Jake Lloyd	Männlich	05.03.1989	10
Dat28	00:58					
Dat29	00:12	E.T. - Der Außerirdische	Henry Thomas	Männlich	09.09.1971	11
Dat30	00:21					
Dat31	00:20	Gossip Girl Staffel 4.09	Taylor Momsen	Weiblich	26.07.1993	17
Dat32	00:19					
Dat33	00:15	Pretty Little Liars 2.02	Sasha Pieterse	Weiblich	17.02.1996	15
Dat34	00:37	Pretty Little Liars 2.04				

Dat35	00:14	Die Legende von Aang	Nicola Peltz	Weiblich	09.01.1995	15
Dat36	00:17					
Dat37	00:24	True Grit	Hailee Steinfeld	Weiblich	11.12.1996	14
Dat38	00:29					
Dat39	00:21	Game of Thrones 3.02	Sophie Turner	Weiblich	21.02.1996	18
Dat40	00:24					
Dat41	00:11	Hotel Zack & Cody 1.01	Cole Sprouse	Männlich	04.08.1992	13
Dat42	00:16					
Dat43	00:12	Hotel Zack & Cody 1.01	Dylan Sprouse	Männlich	04.08.1992	13
Dat44	00:09					
Dat45	00:18	Harry Potter und der Halbblutprinz	Daniel Radcliffe	Männlich	23.07.1989	19
Dat46	00:22					
Dat47	00:29	Transformers - Ära des Untergangs	Nicola Peltz	Weiblich	09.01.1995	19
Dat48	00:20					
Dat49	00:13	Kick-Ass 2	Chloe Moretz	Weiblich	10.02.1997	16
Dat50	00:39					
Dat51	00:27	Insel der Besonderen Kinder	Asa Butterfield	Männlich	01.04.1997	19
Dat52	00:20					
Dat53	00:19	Into the Badlands 2.01	Aramis Knight	Männlich	03.10.1999	18
Dat54	00:34	Into the Badlands 2.02				

Tabelle 1 Übersicht der selbsterstellten Videodateien aus Film- und Serienmaterial, deren Länge, sowie Namen, Geburtstage und Alter zum Drehzeitpunkt der jeweiligen Schauspieler(-innen) (Zahlen hinter Serientiteln bezeichnen Staffel und Episode) , eigene Quelle

Tabelle 2 stellt eine Zusammenfassung über Geschlecht und Alter des Schauspieler zum Drehzeitpunkt dar und berücksichtigt eine Unterteilung in drei Altersgruppen: Kind, Jugendlich und Erwachsen.

	Männlich	**Weiblich**
Gesamt	30	24
Kind	20	6
Jugendlich	4	14
Erwachsen	6	4

Tabelle 2 Übersicht Anzahl an Schauspielern pro Geschlecht und je Altersgruppe, eigene Quelle

Von den insgesamt 30 männlichen Personen sind 20 Kinder, also im Alter von Null bis 13 Jahren, vier Jugendliche im Alter zwischen 14 und 17 Jahren sowie sechs Erwachsene mit einem Alter von 18 Jahren oder älter.

Bei den insgesamt 24 weiblichen Testsubjekten sind nur sechs im Kindesalter zu finden, dafür 14 im Jugendalter. Außerdem sind vier junge Frauen Teil der Untersuchung.

3.2 Verarbeitung der Testdateien

Die Videos werden im ersten Schritt zusammen in einem Ordner gesammelt, der anschließend mit dem Programm FTK-Imager von AccessData Group Inc. in ein Image-Format (.dd oder .e01) gebracht wird, das von XWF gelesen und sicher geöffnet werden kann.

Dies geschieht um einen realistischen Ablauf zu simulieren, wie er bei der forensischen Beweisaufnahme und Datensicherung vorkommt, da hier stets Sicherheitskopien des zu untersuchenden Systems angefertigt werden. Prinzipiell ist auch das reine Hinzufügen des Verzeichnisses, das die Videos enthält möglich.

Das IT-forensische Analyseprogramm öffnet Dateien, Snapshots oder ganze Images in einer schreibgeschützten Umgebung. Hiermit wird sichergestellt, dass die den Fall bearbeitenden Ermittler keinerlei Daten verändern und die Datenintegrität weder absichtlich noch aus Versehen verletzen können. Sämtliche Dateien können byteweise untersucht oder durch spezielle Viewer-Programme geöffnet und betrachtet werden, Veränderungen sind jedoch nicht vornehmbar. Meist ist das digitale Abbild eines Rechners vorliegend, das heißt, es wurde ein Image des logischen oder physikalischen Datenträgers angefertigt.

Auf dem eigenen Computer ist dieses Image lediglich eine einzelne Datei, beinhaltet jedoch oft tausende Dateien. Darunter befinden sich die für die Analyse relevanten Bild- und Videoformate. Der eigene Rechner oder Datei-Explorer kann diese jedoch weder finden noch öffnen, ohne sie zu extrahieren. XWF bietet diese Export-Funktion an und sie wird deshalb der erste Schritt des Workflows, nachdem grundlegende Analyseschritte wie die Erweiterung des Dateiüberblicks und die Dateiheader-Signatur-Suche abgeschlossen sind und nach den Videodateien gefiltert wurde.

Im nächsten Schritt werden die einzelnen Dateien nacheinander dem Videoanalyseprogramm übergeben und von ihm analysiert, um einen möglichen Mehrwert an Informationen zu erzielen, sowie genauere Ergebnisse als eine manuelle Bearbeitung zu liefern.

Dabei wird darauf geachtet inwiefern sich die Alters- und Geschlechtsbestimmungen mit jeder Sekunde verändern. Es wird ein Wert ausgewählt der dem gerundeten Mittel gleichkommt und versucht Fehlberechnungen wie Verschätzungen von groben Fehlern, deren Ursprung die schlichte Nichterkennung der Personen ist, zu differenzieren. Die folgende Übersicht soll die Prozesse nochmal visuell verdeutlichen.

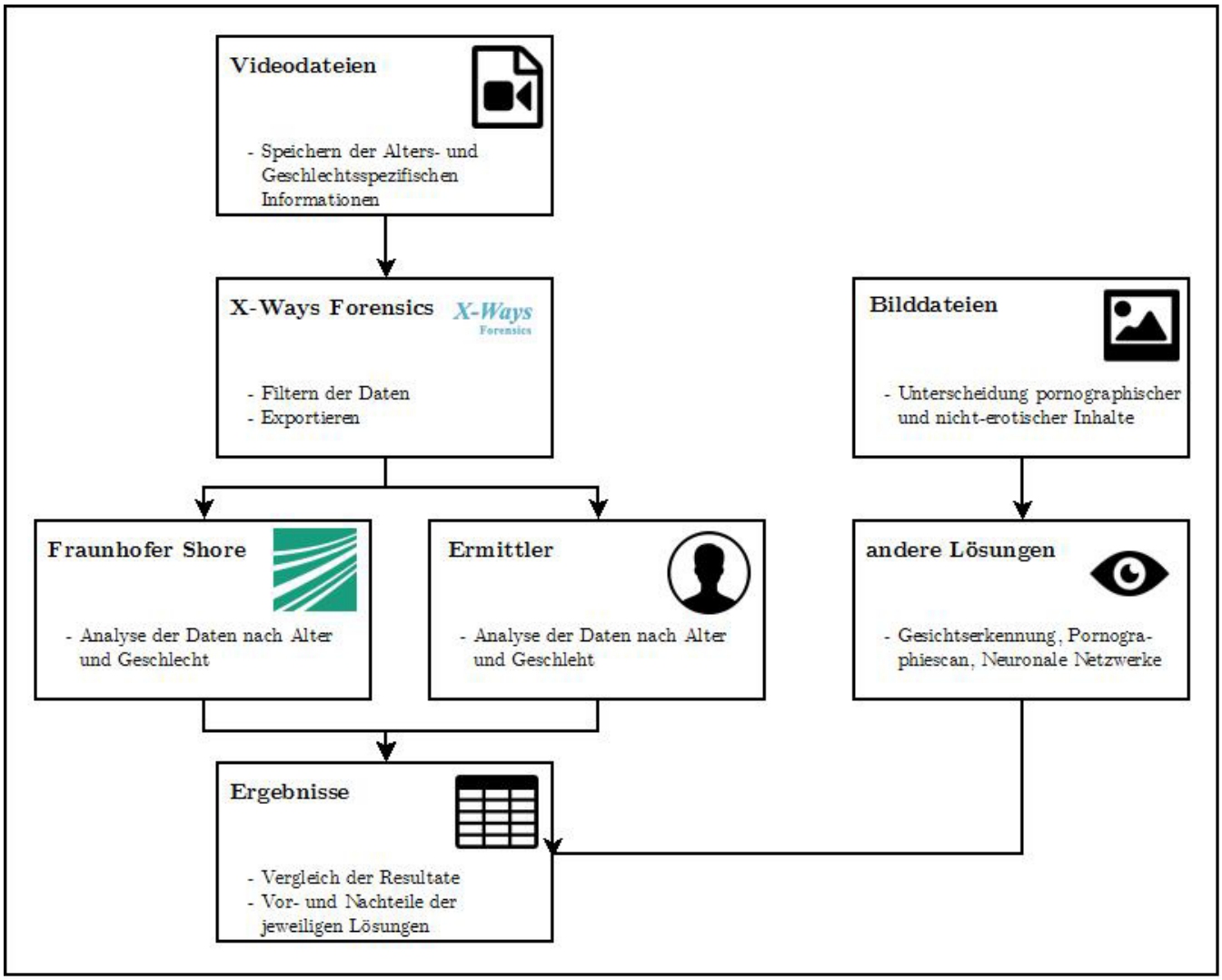


Abb. 10 Workflow der Arbeit, eigene Quelle, Bilder heruntergeladen von: https://cdn2.iconfinder.com/data/icons /font-awesome/1792/file-video-o-128.png, https://forenzniprodukty.cz/wp-content/uploads/2016/08/X-Ways-Forensics-350x257.jpg, https://pbs.twimg.com/profile_images/667367596120014852/LNeIU3mB.jpg, https://cdn2 .iconfinder.com/data/icons/ios-7-icons/50/user_male2-128.png, https://cdn3.iconfinder.com/data/icons/medcare /512/table-512.png, https://cdn0. iconfinder.com/data/icons/faticons-2/28/image33-128.png, https://cdn4 .iconfinder.com/data/icons/ionicons/512/icon-eye-128.png,

Eine Auswertung, der von SHORE gewonnenen Informationen erfolgt im nächsten Kapitel. Screenshots der Gesichtserkennungen mitsamt dem jeweils bestimmten Alter und geschätzten Geschlecht können im Anhang eingesehen werden. Schlussendlich werden selbige Dateien auch in Kapitel 5 zum Vergleich mit anderer Software und manueller Untersuchung genutzt.

4 Ergebnisse

4.1 Geschlechtsbestimmung

Die Analyse durch die Videosoftware war in insgesamt 53 Fällen erfolgreich, nur eine Datei, Dat20, konnte nicht erkannt werden.

Deshalb wird die anschließende Auswertung der Geschlechtsbestimmung und auch der Altersschätzung an 29 Dateien männlicher Schauspieler und 24 Dateien weiblicher Schauspieler durchgeführt. In Tabelle 3 ist die Bestimmung des jeweiligen Geschlechts aufgelistet.

Geschlecht	**Erkannt Männlich**	**Erkannt Weiblich**
Männlich (29)	17	12
Weiblich (24)	0	24

Tabelle 3 Anzahl männlicher und weiblicher Teilnehmer und die Verteilung der Geschlechtsbestimmung durch die Software, eigene Quelle

Von den 29 männlichen Tests wurden 17 als männlich erkannt, jedoch zwölf als weiblich. Die Genauigkeitsrate beträgt 58,6%, und befindet sich damit objektiv gesehen bei einem sehr geringen Wert. Elf der 17 männlichen Kinder (64,7%) und ein Drittel der männlichen Jugendlichen wurden falsch klassifiziert. Die männlichen Erwachsenen hingegen sind richtig eingestuft wurden.

Es wird davon ausgegangen, dass die Einstufung der männlichen Schauspieler als Weiblich vor allem im Kindesalter erfolgte, da in diesem Alter relativ weiche, kindliche Züge, runde Köpfe und große Augen vorliegen. Dies sind erfahrungsgemäß vorrangig weibliche Eigenschaften. Mit zunehmendem Alter werden die Züge der Jugendlichen definierter und härter, Gesichter neigen zur elliptischen Form. Hohe Wangenknochen und zusätzliche Behaarung im Mundbereich und bei den Augenbrauen ergeben harte Kanten. Daher wird die Erkennung im Jugendalter genauer, eindeutig jedoch erst richtig präzise im Erwachsenenstadium.

Im Vergleich dazu wurden alle 24 von 24 Dateien der weiblichen Schauspieler dem richtigen Geschlecht zugeordnet und erzielen damit eine Genauigkeit von 100%. Es gibt keine Abhängigkeit vom Alter.

Die Zuordnung zum weiblichen Geschlecht ist deutlich überklassifiziert und erkennt auch junge männliche Schauspieler, hat im Test jedoch keinerlei Fehler bei den weiblichen Schauspielern erbracht. Die Zuordnung zum männlichen Geschlecht erfolgt unterklassifiziert und erkennt demnach eigentlich männliche Subjekte nicht, jedoch wurde keine Datei mit weiblichem Schauspieler fälschlicherweise erkannt.

Es existiert eine große Diskrepanz zwischen den Resultaten der jeweiligen Bestimmungen abhängig vom Geschlecht. Außerdem hat im Test das Alter der männlichen Subjekte die Geschlechtsbestimmung beeinflusst.

4.2 Altersbestimmung

Nachfolgend wird in Tabelle 4 die Altersbestimmung unabhängig von der jeweiligen Altersgruppe betrachtet, um allgemeine Tendenzen der Analysesoftware zu erkennen. Festzustellen ist, ob diese Tendenzen, wenn vorhanden, in Abhängigkeit vom Geschlecht variieren.

Geschlecht	**Alter genau geschätzt**	**Alter zu hoch geschätzt**	**Alter zu gering geschätzt**
Männlich(29)	1	12	16
Weiblich (24)	3	18	3

Tabelle 4 Geschlechterabhängige Altersbestimmung unterteilt nach exakter, zu hoher oder zu niedriger Altersschätzung, eigene Quelle

Vom den 29 Dateien mit männlichen Schauspielern wurde genau eine mit dem exakten Alter bestimmt, das ergibt eine Genauigkeit von 3,4%. Des Weiteren wurden zwölf der Subjekte älter eingeschätzt als sie wirklich sind, 16 Subjekte jünger als sie wirklich sind. Der Anteil der zu alt und zu jung eingeschätzten Schauspieler liegt bei 55,2% respektive 41,4%.

Es ist zu erkennen, dass hier kaum eine Tendenz festzustellen ist, da das System je einen Großteil der Dateien altern oder verjüngen lässt. Der Anteil der zu jung eingeschätzten Subjekte ist jedoch geringfügig größer und damit existiert eine beinahe marginale Tendenz zur Verjüngung. Dies birgt einen hohen kritischen Fehler.

Bei den Weiblichen Subjekten wurden drei Alter exakt bestimmt, was eine Genauigkeit von 12,5% ergibt. Weiterhin wurden 18 der 24 Teilnehmer als zu alt eingeschätzt. Dies betrifft genau 75% der Personen. Die übrigen drei weiblichen Subjekte wurden jünger eingeschätzt als sie sind, und ergeben einen Anteil von 12,5%.

Bei den weiblichen Tests ist eine deutliche Tendenz der Alterung zu erkennen. Dreiviertel der weiblichen Schauspieler werden älter eingeschätzt als sie es wirklich sind. Auch die beiden größten Ausreißer in den Daten, von neun und zwölf Jahren Altersdifferenz zwischen wahrem und geschätztem Alter, sind im Bereich der weiblichen Jugendlichen zu finden.

Prinzipiell ließe sich sagen, dass die Altersbestimmung sehr ungenau arbeite, da nur ein extrem geringer Anteil exakt eingeschätzt wurde. Es ist jedoch zu vermerken, dass es sich bei diesem Versuch um eine Schätzung handelt, deren Anspruch es ist, einen Wert möglichst ähnlich dem Vorgegebenen zu liefern. Die obige Auswertung in Tabelle 4 betrachtet jedoch nicht wie ähnlich die Resultate sind, sondern stellt nur absolute Übereinstimmung fest.

Es erfolgt keine Unterscheidung zwischen einem Verschätzen von einem Jahr oder zehn Jahren.

Außerdem ist zu bedenken, dass das exakte Alter der Personen nicht unbedingt eine Rolle spielen muss. Es ist für das Projekt schlichtweg nicht relevant ob eine zu identifizierende Person zehn oder elf Jahre alt ist, da die Zuordnung in beiden Fällen zur selben Altersgruppe (Kinder) erfolgt.

Entscheidend ist die exakte Bestimmung jedoch an den Altersgrenzen zwischen den Gruppen, da die Zuordnung einer 13-jährigen Person in die Gruppe der Kinder fällt, während eine 14-jährige der Gruppe der Jugendlichen angefügt wird.

Es wird also eine Untersuchung der einzelnen Altersgruppen abhängig vom Geschlecht benötigt.

In der Statistik werden die beiden Maße Präzision ("Precision") und Sensitivität ("Recall") verwendet um einfache Beurteilungen binärer Klassifikatoren vorzunehmen. Die Sensitivität "gibt den Anteil der korrekt als positiv klassifizierten Objekte an der Gesamtheit der tatsächlich positiven Objekte an"[6]. Sie berechnet sich aus richtig-positiven Ergebnissen geteilt durch richtig-positiven addiert mit falsch-negativen Ergebnissen. Die Präzision "gibt den Anteil der korrekt als positiv klassifizierten Ergebnisse an der Gesamtheit der als positiv klassifizierten Ergebnisse an"[7]. Sie berechnet sich aus richtig-positiven Ergebnissen geteilt durch richtig-positiven addiert mit falsch-positiven Ergebnissen.

Beide Maße werden besonders in der Computerwissenschaft genutzt, sind jedoch stark versuchsabhängig. Außerdem müssen die Aussagen über ihre Werte stets mit Vorsicht getroffen werden, da sie einander stark beeinflussen. Ein Versuch kann zum Beispiel eine sehr hohe Präzision ausweisen, wenn alle negativen Objekte richtig-negativ erkannt werden. Wie viele positive Objekte erkannt werden, spielt dann keine Rolle mehr, weil als Ergebnis immer Eins herauskommt.

Um beide Maße miteinander zu nutzen und eine Beurteilung anhand eines einzigen Wertes durchführen zu können, wurde das F_1-Maß, das gewichtete harmonische Mittel, hinzugezogen. Es wird berechnet aus zwei Mal der Multiplikation von Präzision und Sensitivität geteilt durch die Addition der beiden.

Alle drei Maße sind definiert für Werte zwischen null und eins, und umso höher der Wert des Maßes, umso höher ist die Trefferquote des Systems. Es ist jedoch zu beachten, dass eine Einschätzung des Wertes oft subjektiv ist, und natürlich nicht nur individuell verwertbar aussehen muss, sondern sich auch im Vergleich mit anderen Systemen behaupten sollte.

---

[6] Wikipedia, "Beurteilung eines binären Klassifikators - Wikipedia" https://de.wikipedia.org/wiki/Beurteilung_eines_bin%C3%A4ren_Klassifikators, angesehen am 26. Juni 2017

[7] Wikipedia, "Beurteilung eines binären Klassifikators - Wikipedia" https://de.wikipedia.org/wiki/Beurteilung_eines_bin%C3%A4ren_Klassifikators, angesehen am 26. Juni 2017

Um aber überhaupt von einem "guten" oder "schlechten" Ergebnis sprechen zu können, wird subjektiv der Wert von 0,8 als Grenzwert betrachtet, das heißt alle Versuche, die mehr als Dreiviertel der Personen richtig zugeordnet haben, werden als "gut" angenommen.

Abschließend ist zu bedenken, dass die folgenden Auswertungen keinerlei fallrelevante Unterschiede zwischen zu alt eingeschätzten und zu jung eingeschätzten Personen machen. Für ein gerichtsfestes Gutachten muss jedoch darauf geachtet werden, dass sowohl unkritische als auch kritische Fehleinschätzungen getroffen werden. Unkritisch sind all jene Personen, die in eine höhere Altersgruppe eingeordnet werden; also Kinder, die den Erwachsenen oder Jugendlichen sowie Jugendliche, die den Erwachsenen zugeordnet werden. Hierbei werden eventuell fallrelevante Daten entkräftet, was prinzipiell zu vermeiden ist, aber vor Gericht ist der Ausschluss von zweifelhaften Bildern gängig, um zu verhindern, dass jemand zu Unrecht angeklagt und verurteilt wird. Kritisch hingegen wäre einen Jugendlichen als Kind oder einen Erwachsenen als Kind oder Jugendlichen zu klassifizieren. Dies sind eindeutige Falschaussagen, die die Glaubwürdigkeit des Gutachtens oder Gutachters degradieren. Es ist also besser ein Bild mit kinder- oder jugendpornografischem Inhalt als Erwachsenenpornografie zu erkennen, als andersherum.

Es folgt die Untersuchung der einzelnen Altersgruppen und die jeweilige Berechnung der drei Maße.

Männl. Kinder	**Erkannt**	**Nicht Erkannt**
Sind (20)	16	4
Sind nicht (9)	3	6

Tabelle 5 Erkennung der männlichen Kinder, eigene Quelle

Von den 29 Männlichen Personen sind 20 zwischen null und 13 Jahre alt. 16 dieser Kinder wurden vom Videoprogramm als ebensolche erkannt (richtig-positiv), was eine Genauigkeit von 80% ausmacht. Die verbliebenen vier wurden um mindestens so viele Jahre älter geschätzt, wie für ein Einordnen in eine höhere Altersgruppe nötig war. Diese bezeichnet man als falsch-negativ Resultate oder Fehler erster Art. Es handelt sich hierbei um unkritische Fehler.

Von den neun männlichen "nicht-Kindern" wurden sechs richtigerweise nicht erkannt (richtig-negativ), jedoch drei fälschlicherweise den Kindern zugeordnet. Diese 33% bezeichnet man als falsch-positive Ergebnisse, Fehler zweiter Art oder kritischer Fehler.

Maß	**Ergebnis**	**Berechnung**
Precision	0,842105263	16/(16+3)
Recall	0,8	16/(16+4)
F_1-Maß	0,820512821	2*0,84*0,8/(0,8+0,84)

Tabelle 6 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der männlichen Kinder, eigene Quelle

Die Präzision berechnet sich aus 16/(16+3) und ergibt 0,84, die Sensitivität errechnet sich aus 16/(16+4) und ergibt 0,8. Das F_1-Maß wird kalkuliert durch 2*0,84*0,8/(0,8+0,84) und ergibt 0,82.

Alle drei Werte sind subjektiv gesehen recht gut, jedoch verbesserbar. Eine genaue Einschätzung wird im folgenden Kapitel beim Vergleich mit anderer Software und mit manueller Altersschätzung erbracht.

Männl. Jugendliche	**Erkannt**	**Nicht Erkannt**
Sind (3)	0	3
Sind nicht (26)	4	22

Tabelle 7 Erkennung der männlichen Jugendlichen, eigene Quelle

Von den nur drei Jugendlichen männlichen Schauspielern im Alter zwischen 14 und 17 Jahren wurde kein einziger richtig erkannt, alle drei wurden jünger geschätzt und somit der Altersgruppe der Kinder zugeordnet. Es ist ein deutlicher kritischer Fehler auszumachen.

Zusätzlich wurden vier Personen der Gruppe der männlichen Kinder mit insgesamt 22 Personen als Jugendliche identifiziert. Dies macht einen Anteil von 13,6% aus.

Maß	**Ergebnis**	**Berechnung**
Precision	0	0/(0+4)
Recall	0	0/(0+3)
F_1-Maß	0	2*0*0/(0+0)

Tabelle 8 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der männlichen Jugendlichen, eigene Quelle

Die Präzision berechnet sich aus 0/(0+4) und ergibt 0, die Sensitivität errechnet sich aus 0/(0+3) und ergibt 0. Das F_1-Maß wird kalkuliert durch 2*0*0/(0+0) und ergibt ebenfalls 0. Da nicht ein Subjekt richtig eingeschätzt wurde, sind die Werte aller drei Maße gleich null. Dies ist ein extrem schlechtes Ergebnis und zeigt einen deutlichen Problemherd auf, der jedoch teilweise durch den Mangel an Massendaten verursacht wurde. Der in Kapitel 5.3 folgende Ausblick wird darauf eingehen.

Männl. Erwachsene	**Erkannt**	**Nicht Erkannt**
Sind (6)	6	0
Sind Nicht (23)	0	23

Tabelle 9 Erkennung der männlichen Erwachsenen, eigene Quelle

Alle sechs männlichen Erwachsenen wurden richtig klassifiziert, damit liegt die Trefferrate bei 100%. Die zuvor erwähnte geringe Tendenz zur Verjüngung konnte hier nicht festgestellt werden.

Kein männliches Kind oder Jugendlicher wurde der Gruppe der Erwachsenen zugeordnet, daher sind sowohl der Fehler erster Art als auch der Fehler zweiter Art non-existent.

Maß	**Ergebnis**	**Berechnung**
Precision	1	6/(6+0)
Recall	1	6/(6+0)
F_1-Maß	1	2*1*1/(1+1)

Tabelle 10 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der männlichen Erwachsenen, eigene Quelle

Die Präzision berechnet sich aus 6/(6+0) und ergibt 1, die Sensitivität errechnet sich aus 6/(6+0) und ergibt 1. Das F_1-Maß wird kalkuliert durch 2*1*1/(1+1) und ergibt ebenso 1. Dieses Ergebnis ist absolut perfekt und kann nicht gesteigert werden. Auch hier ist jedoch zu bedenken, dass es sich um eine geringe Menge an untersuchten Daten handelt und die nach oben offene Altersspanne der erwachsenen Altersgruppe eine zu hohe Einstufung des Alters nicht zulässt.

Weibl. Kinder	Erkannt	Nicht Erkannt
Sind (6)	6	0
Sind nicht (18)	2	16

Tabelle 11 Erkennung der weiblichen Kinder, eigene Quelle

Von den sechs weiblichen Kindern wurden alle sechs richtig eingestuft, was eine Genauigkeit von 100% ergibt. Zusätzlich wurden zwei der insgesamt 18 "Nicht-Kinder" als Kinder klassifiziert, was einen Anteil von 11,1% ausmacht.

Maß	Ergebnis	Berechnung
Precision	0,75	6/(6+2)
Recall	1	6/(6+0)
F_1-Maß	0,857142857	2*1*0,75/(1+0,75)

Tabelle 12 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der weiblichen Kinder, eigene Quelle

Die Präzision berechnet sich aus 6/(6+2)und ergibt 0,75, die Sensitivität errechnet sich aus 6/(6+0) und ergibt 1. Das F_1-Maß wird kalkuliert durch 2*1*0,75/(1+0,75) und ergibt 0,86. An diesem Test ist zu erkennen, wie ein extrem hoher Recall-Wert von genau Eins falsche Rückschlüsse mit sich führen könnte, wenn er alleinstehend betrachtet wird. Zwar erkennt die Analyse alle Gesuchten Personen richtig, liefert aber durch eine Überklassifizierung auch Ergebnisse, die nicht dazu gehören. Es existiert demnach kein Fehler erster Art, jedoch ein Fehler zweiter Art. Der Wert des F_1-Maßes beläuft sich in ähnlichem Bereich wie schon zuvor bei den männlichen Kindern.

Weibl. Jugendliche	Erkannt	Nicht Erkannt
Sind (14)	1	13
Sind nicht (10)	0	10

Tabelle 13 Erkennung der weiblichen Jugendlichen, eigene Quelle

Eine extrem niedrige Erkennung hat bei den weiblichen Jugendlichen stattgefunden, hier wurde nur eine einzige Person der richtigen Altersgruppe zugeordnet. 13 der 14 weiblichen Subjekte, und damit 92,9% wurden falsch klassifiziert. Der Großteil wurde zur Gruppe der Erwachsenen hinzugefügt, was einen großen unkritischen Fehler hervorruft.

Es ist also ein sehr hoher Fehler erster Art feststellbar, während hingegen kein Fehler zweiter Art vorhanden ist, da kein weibliches Kind oder eine Erwachsene den Jugendlichen zugeordnet wurde.

Maß	**Ergebnis**	**Berechnung**
Precision	1	1/(1+0)
Recall	0,071428571	1/(1+13)
F_1-Maß	0,133333333	2*1*0,071/(1+0,071)

Tabelle 14 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der weiblichen Jugendlichen, eigene Quelle

Die Präzision berechnet sich aus 1/(1+0) und ergibt 1, die Sensitivität errechnet sich aus 1/(1+13) und ergibt 0,07. Das F_1-Maß wird kalkuliert durch 2*1*0,071/(1+0,071) und ergibt 0,13. Ohne den Fehler der zweiten Art springt der Wert der Precision auf Eins, was alleinstehend bei einem solch schlechten Ergebnis jedoch fatal zu interpretieren wäre.

Dank des extrem niedrigen Recalls, beeinflusst durch den hohen Fehler erster Art, wird das mittelnde F_1-Maß aussagekräftig auf 0,13 gesenkt, einen stark verbesserungswürdigen Wert.

Es ist festzuhalten, dass trotz der mehr als vierfachen Menge an zu untersuchenden Dateien (im Vergleich zu den männlichen Jugendlichen) erneut ein sehr schlechtes Ergebnis bei den Jugendlichen aufgetreten ist. Bei den weiblichen Subjekten ist jedoch die starke Tendenz zur Alterung erkennbar, die dafür sorgt, dass fast ein Dutzend weibliche Jugendliche als weibliche Erwachsene erkannt werden.

Es wird vorläufig festgehalten, dass das Alter der Jugendlichen (zwischen 14 und 17 Jahren) eine Problemquelle sowohl für weibliche als auch für männliche Individuen darstellt.

Weibl. Erwachsene	**Erkannt**	**Nicht Erkannt**
Sind (4)	4	0
Sind Nicht (20)	11	9

Tabelle 15 Erkennung der weiblichen Erwachsenen, eigene Quelle

Die vier weiblichen Erwachsenen sind alle richtig klassifiziert wurden, allerdings wurden weitere elf vermeintliche Treffer durch zu alt eingeschätzte Jugendliche erzielt. Diese elf Personen machen einen Anteil von 55% der weiblichen Kinder und Jugendlichen aus.

Maß	**Ergebnis**	**Berechnung**
Precision	0,266666667	4/(4+11)
Recall	1	4/(4+0)
F_1-Maß	0,421052623	2*1*0,267/(1+0,267)

Tabelle 16 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der weiblichen Erwachsenen, eigene Quelle

Die Präzision berechnet sich aus 4/(4+11) und ergibt 0,27, die Sensitivität errechnet sich aus 4/(4+0) und ergibt 1. Das F_1-Maß wird kalkuliert durch 2*1*0,267/(1+0,267)) und ergibt 0,42. Auch hier sind die Auswirkungen der stetigen Überschätzung des Alters der weiblichen Subjekte deutlich. Es ist fraglich ob die häufige und umfassende Verwendung von Make-Up-Produkten dazu führt, dass das Programm junge Mädchen als erwachsene Frauen ansieht oder der Algorithmus selbst Verbesserungen benötigt. Nichtsdestotrotz liegt ein F_1-Wert von 0,42 weit unter dem anvisierten Ziel.

5 Diskussion

5.1 Vergleich mit anderer Software

Die Recherche zu Softwarelösungen oder softwaregestützten Problemlösungsansätzen, die durch automatisierte Prozesse den manuellen Anteil an Arbeit ersetzen oder zumindest einschränken und damit verkürzen sollen, hat sich durchaus als schwierig herausgestellt, denn die Liste der potentiellen Vorgehensweisen ist kurz. Das manuelle Durchsehen der eventuell inkriminierten Bilder ist bislang das Mittel der Wahl, so aufwendig und belastend es auch sein mag. Gerade deswegen haben jedoch einige Unternehmen, Strafermittlungsbehörden und öffentliche Institute an Lösungen dieser Problemstellung gearbeitet. Der erste Teil von Kapitel 5 setzt sich mit den verschiedenen Lösungsansätzen auseinander, von denen jedoch keiner völlig ausgereift ist oder in großem Ausmaß eingesetzt wird.

5.1.1 Gesichtserkennungssoftware

Neben der in den vergangenen Kapiteln beschriebenen Software existieren weltweit dutzende Möglichkeiten der Implementierung von automatischen Gesichtserkennungsalgorithmen in Softwaresuites. Schon bei einer kurzen Internetrecherche tauchen Namen von OpenSource-Varianten wie OpenFace oder OpenBR (beide stark an OpenCV angelehnt) sowie kommerzielle Lösungen wie Face++, Betaface und Kairos auf. Desweiteren stehen internationale Firmengiganten wie Google oder Microsoft bei einem so hochmodernen und technologieforderten Thema nicht hinten an; hier werden große Ressourcen genutzt um möglichst markbeherrschende Produkte zu vertreiben.

Google besitzt zwei verschiedene Lösungen der Gesichtserkennung: Cloud Vision und Mobile Vision. Ersteres ist eine cloudbasierte Erkennungsoftware, die sich weniger auf die reine Gesichtserkennung und die Bestimmung von Alter und Geschlecht fokussiert, sondern allgemein zu Bilderkennung genutzt wird. Google extrahiert Symbole und Logos, Schriftzeichen und -züge sowie Personen aus dem Bild und verknüpft sie mit möglichen Suchbegriffen, die das Bild bei einer Suchanfrage aufbringen könnten.

Die App Mobile Vision ist eher auf Gesichtserkennung ausgerichtet und kann in Fotos, Videos und Livevideos Gesichtslandmarken erkennen und Emotionen erkennen. Allerdings findet keine Alters- oder Geschlechtsbestimmung statt.

Microsoft's Cognitive Services stellen eine ganze Palette von unterschiedlichen Erkennungssoftwares zur Verfügung wie die Emotionserkennung, die Gesichtserkennung, die Videoerkennung und die Bewegungserkennung.

Es ist ebenso möglich durch eine Kombination dieser Tools ein Produkt zu erschaffen, das mehrere Analysen miteinander verknüpft und umso mehr Informationen gewinnen kann.

Die kostenlosen Test-Schlüssel sind auf je 30 Tage limitiert (90 Tage bei der Videoerkennung), ebenso ist die Transaktionsanzahl begrenzt (zwischen 5.000 und 30.000 insgesamt und je 20 pro Minute) und es existiert ein Dateilimit für Videoübertragungen. Videos selbst können nicht analysiert werden, nur Bilder oder der Livefeed der Laptopkamera.

Die Microsoft-Lösung "VideoFrameAnalysis"[8] zeigt die einzelnen Erkennungssoftwares in Aktion. Dabei wird deutlich, dass pro ausgeführter Erkennung eine Transaktion durchgeführt wird. Wird hierbei die Videokamera des Laptops benutzt, findet eine Analyse des Videoframes mitsamt Alters-, Geschlechts- und Emotionsbestimmung statt. Es werden demnach zeitgleich Transaktionen aller drei APIs ausgeführt.

Microsoft stellt ein flexibles Bezahlmodel zur Verfügung, bei dem je nach Anzahl der benötigten Transaktionen variable Kosten anfallen. Ein derartiges Geschäftsmodell ist jedoch nicht im Sinne der meisten Ermittler, da sie zum Teil tausende Dateien überprüfen wollen und damit die Kosten unkalkulierbar werden. Noch dazu wenn mehrere Anwendungen (mindestens Gesichts- und Videoerkennung) bezahlt werden müssen. Ebenso müssten Videodateien in einem zusätzlichen Verarbeitungsschritt zunächst in Frames segmentiert werden. Am ausschlaggebendsten ist jedoch zu bedenken, dass diese Software-Lösung mittels Microsoft's Cloud Azure funktioniert, das heißt, sämtliche Anfragen (selbst die über das SDK) sind cloudbasiert.

---

[8] Microsoft, "Github - Microsoft/CognitiveServices - VideoFrameAnalysis", https://github.com/Microsoft/Cognitive-Samples-VideoFrameAnalysis, angesehen am 4. Juli 2017

Die Algorithmik der Gesichtserkennung liegt nicht lokal beim Nutzer, sondern bei Microsoft Azure. Für die Verwendung im Bereich der Kinder- und Jugendpornografie müssten die inkriminierten Dateien hochgeladen werden. Dies stellt einen Straftatbestand dar. Daher sind die Microsoft Cogntive Services für diesen Einsatzbereich unbrauchbar.

Unter ähnlichen Problemen leiden die meisten Gesichtserkennungsoftwares. Die folgende Tabelle 17 gibt einen Überblick über die im Rahmen dieser Arbeit betrachteten Programme und fasst die Gründe zusammen, aus denen sie sich nicht für den geforderten Einsatz eignen. Es existieren weiterhin dutzende ähnliche Programme, viele konzentrieren sich aber auf die Erkennung von Gesichtern im Allgemeinen (Gesichtsdetektion), extrahieren Gesichtsfeatures und vergleichen anhand deren Gesichter miteinander(vergleichende Gesichtsidentifikation). Außerdem werden vor allem im Bereich der Videoanalyse Facetracker angeboten, also Programme, die die Bewegungen der Gesichter verfolgen. Dabei wird weder das Alter noch das Geschlecht bestimmt.

Viele Programme besitzen trotz guter Alters- und Geschlechtsbestimmung bei Fotodateien keine Möglichkeit zur Analyse von Videodateien. Oft verfügen die grafischen Oberflächen über die Wahl die Videokamera hinzu zuschalten, jedoch wird nicht der Livefeed selbst, sondern nur ein einzelner Snapshot analysiert. Auch bei aufwendiger Videoanalysesoftware wie von Kairos, die umfangreiche Graphen zur Emotionserkennung ausgibt, wird nur eine Altersbestimmung durch Unterscheidung von Altersgruppen getroffen. Diese sind jedoch recht grob gefasst (je circa 20 Jahre) und es ist dadurch nicht möglich zwischen den relevanten Altersgrenzen zu unterscheiden.

Ein weiterer Punkt ist die bereits genannte ungünstige finanzielle Struktur der Produkte. Anstatt das Produkt mit einer einzigen Rechnung zu begleichen, wird dem Benutzer hier eine Rechnung pro Monat und mit anhängiger Höhe des verwendeten Datenumsatzes gestellt. Es ist davon auszugehen, dass dies für viele Unternehmen von Vorteil ist, weil sie so ihre Kunden, die wahrscheinlich von der Funktionsweise und Genauigkeit des Produktes noch nicht überzeugt sind, nicht durch einen hohen Betrag abschrecken wollen.

Anstelle dessen wird angepasst an die benötigte Masse der Benutzungen des Produktes ein monatlicher Betrag vereinbart, der ein kostengünstigeres Produkt suggeriert.

Das aber wohl entschiedenste Ausscheidungskriterium ist die zwanghafte Bindung an eine online arbeitende Lösung. Diese cloudbasierten Services sind für die Verwendung im Bereich der Strafermittlung absolut nicht zu gebrauchen, da sie eine Straftathandlung darstellen.

Programm	**Lokale Verabeitung**	**Einmalige Fixkosten**	**Video-erkennung**	**Altersbestimmung**
	API: Nein, SDK:			
Kairos	Ja	Nein	Ja	Ja, Nicht bei Videos
Betaface	Nein	Nein	Nein	Ja
FaceMark	Nein	Nein	Nein	Nein
FaceRekt	Nein	Nein	Nein	Nein
	API: Nein, SDK:			
EmoVu	Ja	Nein	Ja	Nein
SkyBiometry Face Recognition	Nein	Nein	Nein	Ja
Lambda Labs Face API	Nein	Nein	Nein	Nein
Eyedea EyeFace	Ja	Ja	Ja	Ja
	API: Nein, SDK:			
Face++	Ja	Nein	Nein	Ja
Animetrics FaceR	Nein	Nein	Nein	Nein
Meerkat Facial Recognition	Nein	Nein	Nein	Nein
Heaven on Demand	Nein	Nein	Nein	Nein
Imacondis Face SDK	Nein	Ja	Nein	Nein
Flandmark	Ja	Ja	Ja	Nein
Semantic Vision Technologies	Ja	Ja	Ja	Nein
Cognitec FaceVACS	Ja	Unbekannt	Nein	Ja
Sightcorp F.A.C.E.	Nein	Nein	Nein	Ja
Sightcorp InSight	Ja	Unbekannt	Nein	Nein
Visage Technologies FaceAnalysis	Ja	Unbekannt	Ja	Ja
OpenBR	Ja	Ja	Nein	Ja
OpenFace	Nein	Ja	Nein	Nein

Tabelle 17 Übersicht weitere Software zur Gesichtserkennung, sowohl kommerziell als auch OpenSource, eigene Quelle

Nur die beiden Lösungen von Eyedea und Visage Technologies wären für den gesuchten Einsatz zur Untersuchung von Videodateien auf kinder- und jugendpornografische Inhalte geeignet.

Bei Eyedea Eyefaces GUI ist die Altersanzeige schlecht gelöst, da sie anhand einer vierteiligen Skala (ein Abschritt entspricht 20 Jahren) angegeben wird. Diese Darstellung ist sehr ungenau, noch dazu da der bewegliche Anzeiger bereits so breit ist, wie circa fünf Jahre. Ein Test der SDK könnte hier jedoch Abhilfe schaffen, da die Angaben in einer .json- oder .csv-Datei gespeichert werden. Dazu wäre jedoch der Kauf einer Lizenz nötig, ohne zu wissen inwiefern das Produkt zuverlässig arbeitet.

Die Software von Visage Technologies ist in der Anzahl der zu verarbeitenden Daten begrenzt. Daher ist ein Vergleich der Analyseresultate mit SHORE nicht möglich. Auf eine Anfrage hin, eine erweiterte Demo-Lizenz oder eine SDK-Lösung, zur Übergabe aller Videodateien, und das Bezahlmodell betreffend, erteilte die Visage Technologies seit Wochen keine Antwort.

Der größte Nachteil aller Gesichtserkennungssoftwares inklusive SHORE besteht darin, dass überhaupt Gesichter im Videomaterial vorhanden sein müssen. Ist dies nicht der Fall kann die automatische Bestimmung nicht durchgeführt werden. Bei vielen Produktionen von Kinder- und Jugendpornografie sind keine Gesichter zu erkennen, weil die Gesichter verhüllt, verschleiert oder mit Weichzeichnern verblendet werden, lediglich die Unterkörper der Personen zu erkennen sind, oder es sich nur um Rückansichten der Personen handelt, bei denen nur der Hinterkopf nicht aber das Gesicht eingesehen werden kann. Daher ist zu überlegen ob es nicht sinnvollere Vorgehensweisen gibt, die illegale pornografische Inhalte auch unabhängig vom Gesicht, sondern zum Bespiel anhand der Ausprägung von Beharrung oder dem Bildanteil von Hautfarben erkennen können.

Außerdem besteht die Möglichkeit auf ein Neutrales Netzwerk zurückzugreifen, dass inkriminierte Inhalte nicht erkennt, indem es spezifische Merkmale eines Gesichtes extrahiert, sondern indem es antrainiert wurde Bilder verschiedener Art dank ihrer kombinierten Merkmale zu unterscheiden.

Auf einige dieser Methoden wird folgend eingegangen.

5.1.2 Eingebaute Features von IT-Forensik-Software

X-Ways Forensics und andere Forensiksoftwares besitzen ebenfalls Mittel und Wege, die zur Erkennung von Kinder und Jugendpornografie dienen sollen. XWF liefert dazu sein internes Feature "Hautfarben und Schwarz-Weiß-Erkennung". Prinzipiell liefert die Analyse einen prozentualen Verhältniswert pro Bild und gibt an wie viel des Bildes Hautoberflächen zeigt.

Jedoch kann schon bei einem kurzen Test, bestehend aus fallrelevanten kinderpornografischen Inhalten sowie völlig unbedeutenden Portraits von Schauspielern, erkannt werden, wie extrem ungenau die Analyse vorgeht. Im Test erhielten Bilder beiden Ursprungs mit zwischen 50 und 80 Prozent Hautfarbenanteil nur einen Analysewert zwischen vier und 13 Prozent. Diese viel zu geringe Genauigkeit ist vor Gericht nicht verwertbar.

Des Weiteren ist zu bedenken, dass selbst bei genauer Analyse keinerlei Unterschiede zwischen nicht-pornografischen, legalen pornografischen und illegalen pornografischen Inhalten gemacht wird, da alle hohe Hautfarbenanteile besitzen. Dieser Vorgang kann demnach maximal genutzt werden um Icons vom Betriebssystem oder installierten Programmen sowie irrelevante Landschaftsbilder und Ähnliches auszuschließen.

Die Hautfarbenerkennung funktioniert ausschließlich bei Bildformaten. Das bedeutet für Videodateien, dass sie zuvor in Einzelbilder (Frames) zerlegt werden müssen, was jedoch je nach Anzahl der Videodateien zeitintensiv ist.

Dazu bietet X-Ways ein eigenes Feature an, dass jedoch den MPlayer erfordert, der zwingenderweise im gleichen Verzeichnis wie die zu untersuchenden Daten vorhanden sein muss.

Erfahrungsgemäß ist aufgefallen, dass er nur sehr wenige Formate überhaupt öffnen kann. Aus diesem Grund besitzt die DigiFors GmbH eine hausinterne X-Tension, die auf FFmpeg zurückgreift.

Seit 2008 besitzt das Unternehmen LTU Technologies ein Tool namens LTU Finder, das als externes und einzeln zu erwerbendes Plug-In für Guidance Software's Encase Forensic Sofware verkauft wird.

Im Mai 2015 wurde die Firma vom japanischen Unternehmen JASTEC Co. Ltd gekauft, die seither sämtliche Rechte an LTU Technologies, jetzt JASTEC France, hält. Laut eigenen Angaben verfügt das Tool über erweiterte Bild- und Videoerkennung sowie Identifizierungsmöglichkeiten für Textdateien, die großangelegte Dateisuchen automatisieren. Das Plug-In wird benutzt um Videodateien zu segmentieren und zu extrahieren und um anschließend inkriminierte Dateien automatisch zu erkennen.[9]

Um die Effizienz und Genauigkeit dieses Produktes einschätzen zu können; wurde eine diesbezügliche Onlinerecherche in Gang gesetzt. Festzustellen ist, dass es keinerlei wissenschaftliche Artikel oder Paper gibt, die von der Benutzung des Tools berichten, auch keine Blogeinträge in Forensischen Foren oder Referenzen aus beispielsweise Zeitungsartikeln zu finden sind. Auch eine Nachfrage bei den Ermittlern der DigiFors ergab keinerlei Informationszuwachs.

Eine Googleanfrage mit den Keywords "ltu finder" im zeitlichen Rahmen von 2010 bis 2017 ergab gerade mal vier Ergebnisse, darunter die oben zuvor zitierte Seite. Von den restlichen Einträgen war eine Seite nicht zu erreichen, die zweite listete lediglich dutzende Forensiktools auf, ohne auf ihre Handhabung einzugehen. Die dritte Seite (brasilianischer Herkunft) enthielt eine Präsentation in portugiesischer Sprache, die augenscheinliche auch in die Jahre gekommen war (Hinweise wie Icons veralteter Programme und Geräte, Bezeichnung "Sony Ericsson" (Trennung erfolgte bereits 2012) etc.), sich aber dennoch mit der Herangehensweise von LTU Finder beschäftigte.[10] Den Ausführungen war zu entnehmen, dass LTU Finder die sogenannte Foto-DNA des Bildes mit anderen Bildern vergleicht. Sowohl die Suchmaschinen Bing als auch Yahoo fanden keine weiteren Referenzen zum LTU Finder.

Eine Recherche auf der offiziellen Homepage von LTU Technologies (JASTEC France) ergibt ebenso wenig Aufschlussreiches. Keines der Menüs erhält Informationen zum LTU Finder, lediglich das Programm LTU Engine und die Online-Lösung LTU Cloud werden hier

---

[9] LTU Technologies, "LTU technologies Releases LTU-Finder 3.0 at CEIC",

http://www.ltutech.com/news/ltu-technologies-releases-ltu-finder-3-0-at-ceic/, angesehen am 3. Juli 2017

[10] TechBiz Forense Digital, "ARMAS PARA COMBATE AOS CRIMES DIGITAIS", http://slideplayer.com.br/slide/371771/, angesehen am 3. Juli 2017

vermarktet. Auch unter "About" und "registered trademarks" werden nur LTU, LTU Technologies, LTU Enigine, LTU Cloud und LTU OnDemand gelistet.

Es wird davon ausgegangen, dass die Entwicklung von LTU Finder eingestellt wurde oder nur noch als Teil des Hauptprogrammes LTU Engine erwerblich ist.

Weiterhin ist aus den Demos und Menüs der LTU Engine zu erkennen, dass sich dieses Programm ausschließlich mit dem Vergleich von Bildern beschäftigt.

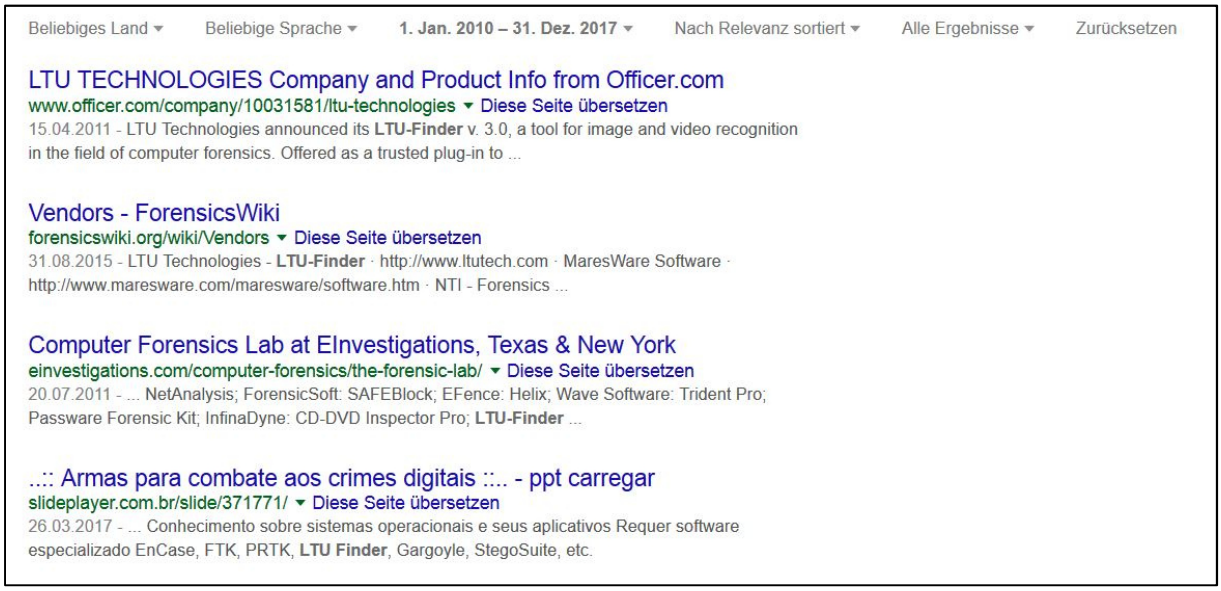


Abb. 11 Google-Ergebnisse für "ltu finder" von 2010 bis 2017, angesehen am 3. Juli 2017 https://www.google.de/search?q=ltu+finder&client=firefox-b-ab&source=lnt&tbs=cdr%3A1%2Ccd_min%3A2010%2Ccd_max%3A2017&tbm=#tbs=cdr:1,cd_min:2010,cd_max:2017&q=%22ltu+finder%22

Die grundlegende Idee des Vergleichens von Inhalten scheint logisch, wenn ein eindeutig inkriminiertes Bild vorliegt, sodass es dann mit allen anderen Bildern abgeglichen werden kann, um Ähnliche zu finden. Jedoch ist zu bedenken, dass viele kinderpornografische Inhalte nicht vom selben Set oder derselben Reihe sind; der Ähnlichkeitsbegriff müsste sehr grob gefasst werden, was zu einer großen Anzahl an falsch-positiven Ergebnissen führt.

Ebenso muss die Eingangsdatei entsprechend definiert sein, und zuvor manuelle gesichtet wurden sein. Abhängig vom Fall scheint dieses Verfahren mehr zusätzlich Arbeit hervorzurufen als zu vermindern.

Kylie Tanner fügt in ihrer Masterarbeit "Modelling Automated Detection of Children in Images" [29 S.8] noch hinzu, dass zum Benutzen des LTU Finders sowohl die Forensiksuite Encase als auch das Plug-In selbst käuflich erworben werden müssen.

Im Gegensatz zum in X-Ways integrierten Tool, bedeutet dies einen finanziellen Mehraufwand, der potentielle Käufer von der Benutzung abhalten könnte. Dies könnte außerdem zu der geringen Resonanz auf das Tool beitragen.

5.1.3 RedLight

Das Digital Forensics and Cyber Security Center der University of Rhode Island besitzt seit 2010 den Pornografie-Scanner "RedLight". Das Programm findet schnell pornografische Inhalte in Bildern und Videos, stellt Verifizierungsmöglichkeiten für den Ermittler durch Thumbnails zur Verfügung und ist in der Lage ausgewählte Dateien, Berichte oder Hash-Sets zu exportieren.[11] RedLight erkennt hohe Konzentrationen an Hautfarben und Kanten menschlicher Formen. Nach eigenen Angaben erfasst es circa 80% richtig, führt zu wenig falsch-positiven Ergebnissen und ist zwischen fünf- und zehnmal zeiteffektiver als vergleichbare Produkte wie LTU. [12]

Besonders vorteilhaft ist die Möglichkeit des Exports von Hash-Sets nach der Verwendung des Programms, da diese von Forensiksuites wie XWF oder Encase importiert und weiterführend analysiert werden können. Nachteilhaft ist die reine Erkennung von Pornografie per se, da keine Unterteilung in legale Erwachsenenpornografie und illegale Kinder - und Jugendpornografie unternommen wird.

Im Januar 2011 haben das National Law Enforcement and Correction Technology Center (NLECTC) und das Electronic Crime Technology Center of Excellence (ECTCoE) des

[11] University of Rhode Island, "RedLight - Research - Digital Forensics and Cyber Security Center at the University of Rhode Island", http://www.dfcsc.uri.edu/research/RedLight, angesehen am 4. Juli 2017

[12] University of Rhode Island, "RedLight Features - Research - Digital Forensics and Cyber Security Center at the University of Rhode Island", http://www.dfcsc.uri.edu/research/RedLightFeatures, angesehen am 4. Juli 2017

National Institute of Justice (NIJ) einen Evaluation Report (Evaluierungsbericht) über die RedLight Software verfasst [30].

Diese war zum damaligen Zeitpunkt in der Beta-Version verfügbar (0.1.0.0). Es wurden insgesamt fünf verschiedene Tests durchgeführt, die in der folgenden Übersicht dargestellt sind.

In Tabelle 18 ist zu erkennen, dass RedLight innerhalb von kurzer Zeit mehrere Zehntausend Bilder analysieren kann und dabei die Anzahl der eventuell inkriminierten auf wenige hundert Bilder senkt.

Der Auswertung des Teams zufolge, handelt sich jedoch bei den gefundenen Resultaten nicht um pornografische Inhalte, sondern lediglich um den Fehler zweiter Art.

Dieser liegt jedoch nur zwischen 1,2 und 2,3 Prozent und ist damit sehr gering. Es existieren keine Angaben über Fehler erster Art. Test 4 zeigt, dass RedLight nicht in der Lage ist E01-Dateien auszuwerten [30].

Test	Umgebung	Dauer der Analyse	Gesamtzahl Bilder	Als Pornografie erkannte Bilder
1	reines, neu aufgesetztes Windows	15 Minuten	13.736	314
2	System von Test 1, Dateiheadersuche	15 Minuten	14.932	359
3	System mit unbekanntem Inhalt	14 Minuten	27.786	342
4	E01-Image des Systems von Test 3	0 Minuten	0	0
5	FTK-Case des Systems von Test 3	12 Minuten	29.721	358

Tabelle 18 Test des NIJ mit jeweiliger Umgebung, Dauer und Resultate der Analyse von RedLight, eigene Quelle

Es ist festzustellen, dass die Anzahl der zu sichtenden Bilder extrem verringert wurde.

Das NIJ stell daher fest, dass zwar noch immer manuelle Arbeit eines Ermittlers erforderlich ist, aber diese drastisch gesenkt werden kann [30]. Über die Eingliederung des Produktes in den Standardablauf der Ermittlungen wird nachgedacht.

Um die Genauigkeit von RedLight selbst zu überprüfen, wurden drei unterschiedliche Tests mit der aktuellen Version 1.2d (Beta) durchgeführt. Test 1 umfasst zehn Portraitbilder erwachsener Schauspieler, die in Tabelle 19 aufgelistet werden.

Datei-name .jpg	**Name Schau-spieler**	**Quelle**	**Erkannt als Pornografie**
Dat1	Robert Downey Jr.	https://s-media-cache-ak0.pinimg.com/originals/da/dc/d4/dadcd453713f86372e4297352939976b.jpg	Ja
Dat2	Kiefer Sutherland	https://image.gala.de/20519312/large1x1-460-460/4243f2a 58b6bfbf09e7f78cc25509967/JN/kiefer-sutherland--10016056-.jpg	Ja
Dat3	Tom Hiddleston	https://upload.wikimedia.org/wikipedia/commons/3/39/Tom_Hiddleston_in_2013.jpg	Ja
Dat4	Brad Pitt	http://img.usmagazine.com/article-leads-vertical-300/1250530894_brad_pitt_290x402.jpg	Ja
Dat5	Sylvester Stallone	https://image.gala.de/20520344/2x3-620-930/e4b46564bdf69a2c9ee0f56d182ee8b7/HT/sylvester-stallone-cm--8555947-.jpg	Ja
Dat6	Angelina Jolie	http://media.vanityfair.com/photos/57e15c417dd0d7d276c7cb7c/master/h_590,c_limit/angelina-jolie-vf-december-2014-ss02.jpg	Ja
Dat7	Helen Mirren	http://thatmomentin.com/wp-content/uploads/2016/05/helen-mirren.jpg	Nein
Dat8	Jenifer Lawrence	http://img.usmagazine.com/article-leads-vertical-300/1298054821_jennifer-lawrence-402.jpg	Ja
Dat9	Nathalie Dormer	https://upload.wikimedia.org/wikipedia/commons/b/b2/Natalie_Dormer_2014.jpg	Ja
Dat10	Robin Wright	https://www.welt.de/img/vermischtes/mobile138426852/1622500487-ci102l-w1024/RTX18SIE-jpg.jpg	Ja

Tabelle 19 Test 1 von RedLight mit zehn Schauspielerportraits, Quellen angesehen am 4. Juli 2017, eigene Quelle

Die grundlegende Problematik von RedLight wird bereits im ersten Test deutlich. Das Programm erkennt sämtliche Dateien bei denen A) der Hautfarbenanteil einem bestimmten Grenzwert überschreitet und/oder B) die Kantendetektion menschliche Formen erkennt, als Pornografie. Neun der zehn Portraits wurden fälschlicherweise als Pornografie erkannt.

Eine Erweiterung des ersten Tests um vier Bilder von Schauspieler Idris Elba (teils mit freiem Oberkörper um den Hautfarbenanteil zu erhöhen) und zwei von Jamie Foxx ergab jedoch ebenso nur neun gefundene Ergebnisse. Die dunkle Hautfarbe der beiden Männer mit afrikanischen Wurzeln wird vom Programm nicht als Hautfarbe erkannt. Es wurden daraufhin zwei Bilder von menschlichen Albinos hinzugefügt; auch sie wurden nicht erkannt.

Die schlechte Erkennung von zu dunkler oder zu heller Hautfarbe könnte je nach Fall einen Fehler erster Art mit sich führen, der zu einem Mehraufwand an Sichtungsmaterial für den Prüfer führt.

In Test 2 werden zehn Bilder mit ausschließlich pornografischem Inhalt verwendet. Hierbei werden alle zehn Dateien richtigerweise als Pornografie identifiziert. Ferner wurde je ein Bild aus Test 1 (nicht-pornografisch) und Test 2 (pornografisch) mittels Schwarz-Weiß-Filter in ein Grautonbild umgewandelt. Beide wurden von RedLight nicht erkannt.

Festzuhalten ist, dass das Programm jedes Bild mit einem ausreichend hohen Anteil an (hellhäutiger) Hautfarbe als Pornografie klassifiziert. Es existiert keine Unterscheidung zwischen nicht-erotischen, legal-erotischen und illegal-erotischen Inhalten, solange genügend Haut zu erkennen ist.

Diese Unterscheidung wäre zwar sinnvoll und nützlich, jedoch ist es auch so möglich das Programm zur Massendatenreduzierung zu nutzen. Wenn bei einem Fall mehrere tausende Bilddateien vorliegen und im ersten Schritt mittels RedLight irrelevante Bilder wie Icons von Betriebssystemen und Programmen sowie Landschaftsbilder und Ähnliches aussortiert werden, muss der Ermittler nur noch die übrigen von RedLight erkannten Bilder manuell auf kinder- und jugendpornografische Inhalte überprüfen.

Dateiname.jpg	Quelle	Erkannt als Pornografie
Dat11	http://208.94.232.100/np/thumbs/uf/332607.jpg	Ja
Dat12	http://208.88.225.212/np/thumbs/rf/329556.jpg	Ja
Dat13	http://199.80.52.156/np/thumbs/Bf/339848.jpg	Ja
Dat14	http://208.94.232.100/np/thumbs/207/207012.jpg	Ja
Dat15	http://208.94.232.100/np/thumbs/Mf/350122.jpg	Ja
Dat16	http://208.88.225.212/np/thumbs/215/215121.jpg	Ja
Dat17	http://199.80.52.149/np/thumbs/df/315729.jpg	Ja
Dat18	http://208.88.225.212/np/thumbs/212/212941.jpg	Ja
Dat19	http://208.88.225.212/np/thumbs/248/248726.jpg	Ja
Dat20	http://199.80.52.156/np/thumbs/If/346838.jpg	Ja

Tabelle 20 Test 2 von RedLight mit zehn pornografischen Bildern, Quellen angesehen am 4. Juli 2017 eigene Quelle

Im dritten Test wird RedLight mit den 54 Videodateien aus Kapitel 4 getestet. Alle 54 Dateien werden erkannt und vom Programm in 1:20 Minuten durchsucht. Dabei werden sechs Dateien als pornografisch erkannt.

Dies sind Dat3, Dat4, Dat6, Dat32, Dat43 und Dat44. Dat3 und Dat4 sowie Dat43 und Dat44 entstammen demselben Ursprung. Eine Wiederholung des Test erbrachte die gleichen Resultate.

Es ist unklar, wieso genau diese Dateien ausgewählt wurden; es wird angenommen, dass die Szenen aus vielen Close-Up-Aufnahmen der Gesichter bestehen, sodass der Hautfarbenanteil höher ist als bei den anderen Videodateien.

Aufgrund der hohen Geschwindigkeit, der einfachen Verwendungsweise und des angeblich Nichtvorhandenseins von Fehlern erster Art wurden von Mitarbeitern der DigiFors GmbH

weitere Tests an Echtfällen vorgenommen, auf die aus sicherheitstechnischen Gründen nur oberflächlich eingegangen werden kann. Es waren deutliche Fehler erster und zweiter Art mit jeweils vierstelligen Ergebnissen zu vermerken. Daher muss nach der Anwendung von RedLight doch jedes einzelne Bild nochmals gesichtet werden, um die Angaben des Programms zu überprüfen. Somit entsteht keine Verkürzung des Arbeitsprozesses. Von der Verwendung von RedLight wird daher abgeraten.

Sowohl Kylie Tanners Masterarbeit [29] als auch Timothy Balls Dissertation [31] widmen sich der Verbesserung von RedLight durch Erstellen eines zusätzlichen Features basierend auf Haar-Klassifikatoren, das Kinder von Jugendlichen und Erwachsenen unterscheiden kann. Beide kommen zu dem Entschluss, dass dies möglich ist. Vor allem Ball erhält hohe Trefferraten bei der Unterscheidung von Kindern unter 12 Jahren. Eine Implementierung dieser Erkennung in RedLight fehlt aktuell jedoch.

5.1.4 Neuronale Netzwerke

Eine in den letzten Jahren häufiger anzutreffende Form der Erkennung inkriminierter Inhalte ist die Verwendung von Neuronalen Netzwerken. Lange Zeit schon genießen sie einen Beliebtheitsgrad durch ihr adaptives Verhalten, konnten aber nie dieselben Trefferquoten aufweisen wie andere Vorgehensweisen. Außerdem ist das Trainieren des Algorithmus sehr aufwendig und stark vom verwendeten Material abhängig. Nichtsdestotrotz könnte die angelernte Maschine aufgrund ihrer über lange Zeit und mit Massendaten angelegten Entscheidungsfähigkeit einen Großteil der manuellen Arbeit eines Ermittlers abnehmen. Folgend werden drei hochmoderne Produkte der letzten Jahre vorgestellt, die sich diesem Ziel gewidmet haben.

Zum Einen arbeitet ein bis zu sechsköpfiges Team am Fraunhofer Institut für Produktionsanlagen und Konstruktionstechnik (IPK) sowohl am Projekt "desCRY" als auch am daraus resultierenden Projekt "TRAFFIIC". Das Ziel von desCRY "ist die Erkennung und Analyse von kinderpornografischen Bild- und Videoinhalten auf sichergestellten

Datenträgern sowie die automatische Erfassung weiterreichender kriminalistischer Ansatzpunkte und die gerichtsverwertbare Dokumentation des Prozesses hierzu."[13]

Dazu verwendet das IPK eine Mustererkennung, die automatisch Inhalte klassifizieren kann und dabei Gesichter und Hautfarbenflächen erkennt. Informationen werden in Datenbanken zwischengelagert um für spätere Anwendungen zur Verfügung zu stehen um Vergleiche zwischen schon erfassten und neu hinzu geführten Inhalten durchführen zu können.

Im Rahmen von TRAFFIIC wurde diese Software durch den österreichischen Projektpartner SEC Technologies um Bilder aus dem Netz sowie generalisierende Klassifikatoren erweitert, die fähig sind, unbekanntes Material zu analysieren.[14]

Das Neuronale Netzwerk wird mit circa 30.000 Dateien gespeist und lernt im ersten Analyseschritt welche davon nicht-erotischen und welche erotischen Inhaltes sind. Im folgenden zweiten Schritt wird nochmals die Unterscheidung der erotischen Inhalte in pornografische und kinderpornografische Bilder antrainiert.

Jede einzelne Klassifikation wird durch die Merkmale der Bilder bestimmt, die ihr zugeordnet werden, das heißt, alle Bilder einer Gruppe besitzen eine große Anzahl gleicher Merkmale mit einer definierten Variationsmöglichkeit. Das Netzwerk erlernt durch Massendaten wie die drei Klassifikatoren beschrieben sind.

Umso mehr Eingangsdaten zur Verfügung stehen, umso exakter ist die Definition jeder einzelnen Klasse. Außerdem steuert der Inhalt der Trainingsdaten die Klassifikation; sie ist gravierend von den Eingangsdaten abhängig und durch weiteres Hinzufügen von Daten adaptiv.

---

[13] Fraunhofer IPK, "Fraunhofer IPK: desCRY", https://www.ipk.fraunhofer.de/projekte/einzelnes-projekt/?tx_ttnews%5Byear%5D=2012&tx_ttnews%5Bmonth%5D=08&tx_ttnews%5Bday%5D=28&tx_ttnews%5Btt_news%5D=84&cHash=d3109c02727f06762e131bba15d9a8f1, angesehen am 3. Juli 2017

[14] Fraunhofer IPK, "Fraunhofer IPK: TRAFFIIC - Traffic analysis for incriminating image content", https://www.ipk.fraunhofer.de/projekte/einzelnes-projekt/?tx_ttnews%5Byear%5D=2016&tx_ttnews%5Bmonth%5D=11&tx_ttnews%5Bday%5D=22&tx_ttnews%5Btt_news%5D=365&cHash=0f7f7c8472f4ccdc331f0e18e23d740d, angesehen am 3. Juli 2017

Schlussendlich kann die angelernte Maschine eine Zuordnung eines neuen Bildes zu einer der drei Kategorien vornehmen, nämlich zu der, mit deren Inhalten die größten Ähnlichkeiten bestehen und die gleichen Merkmale extrahiert werden können.

Jedoch ist den Trefferangaben und Fehlerraten der Fraunhofer-Präsentation (nur den Mitarbeitern der DigiFors GmbH einschließlich des Autors zur Verfügung gestellt) zu entnehmen, dass nur rund 35% der kinderpornografischen Inhalte erkannt werden.

Zudem existieren deutliche Fehler erster und zweiter Art. Dies bedeutet kinderpornografische Bilder werden nicht als solche erkannt und nicht-erotische und legal-erotische Bilder werden als kinderpornografisch erkannt. Letztendlich läuft es darauf hinaus, dass der Ermittler sowohl die als inkriminiert als auch die als unbedeutend gefilterten Bilder manuell sichten muss, um die fehlerhaft unterschiedenen herauszusuchen. Es existiert demnach keinerlei Reduzierung des Arbeitsaufwandes. Auch die weiteren Tests bei der DigiFors GmbH konnten die geringe Genauigkeit und enorm hohe Fehlerraten bestätigen, was zu einer geringen Akzeptanz des Produktes bei den Ermittlern führte. Da das Fraunhofer Institut einem Test von TRAFFIIC im Rahmen dieser Arbeit nicht zugestimmt hat, konnten die zuvor verwendeten Testbilder nicht untersucht werden.

Zum Anderen arbeitet das LKA Niedersachsen seit einiger Zeit an einem ähnlichen Modell zur Bekämpfung und Aufklärung von Verbreitungsdelikten von kinder- und jugend-pornografischen Inhalten. Das Projekt basiert auf der KNIME Analytics Platform, einer Software zur Datenanalyse, -aufbereitung und -bereitstellung. Neben dem Data-Mining von Massendaten ist KNIME in der Lage basierend auf den Daten Vorhersagen zu treffen und Hochrechnungen zu betreiben. Hierbei wird keine lernende Maschine trainiert sondern vielmehr ein vielschichtiger Workflow aus Data-Mining-Schritten erstellt, der Daten durch statistische Analysen klassifiziert oder clustert und dadurch unterteilt.

Zuletzt ist die Lösung des Software-Konzerns Yahoo zu nennen, der das Framework des Neuronalen Netzwerkes "Caffe" als OpenSource etabliert hat. Darauf basierend hat im Jahre 2016 Yahoo das sogenannte "Open NSFW model" ebenso frei über Github zur Verfügung gestellt. Der Slogan "Not Safe For Work" oder "Not Suitable For Work" wird im Internetsprachgebrauch oft für nicht jugendfreie Inhalte verwendet.

Das Neuronale Netzwerk bekommt Bilddateien übergeben und errechnet aufgrund seiner antrainierten Algorithmik einen NSFW-Score zwischen Null und Eins. Umso höher dieser Wert, umso größer ist die Wahrscheinlichkeit dafür, dass es sich bei dem Bild um pornografische Inhalte handelt.

Je nach Anwendung können diese nach einem bestimmten Grenzwert gefiltert werden oder aber als Suchergebnissen auf- oder absteigend angeordnet werden. [15] Yahoo zufolge deutet ein Score von mehr als 0.8 erfahrungsgemäß auf Pornografie hin.

Zur Analyse der Bilder wurden sie auf 256*256 Pixel angepasst, horizontal gespiegelt um die Daten zu vermehren und damit das System stabiler und weniger überangepasst zu gestalten. Anschließend werden die Bilder zufällig in 224x224 Pixelgroße Bilder geschnitten.

Yahoo hat sämtliche Funktionen frei erhältlich veröffentlicht, um interessierten Entwicklern weltweit die Möglichkeit zu geben, das Modell zu testen sowie Feedback für Verbesserungen zu liefern.

Yahoo spricht dabei vom "spirit of collaboration"[16], dem Geiste der Zusammenarbeit. Kurz zuvor (September 2016) jedoch fand der Datendiebstahl von mehr als einer halben Milliarde Yahoo-Nutzern statt, der weltweit für Aufmerksamkeit sorgte und auf den im Dezember ein Zweiter folgte. Unbestätigten Gerüchten zufolge sei bei diesem Hack auch Code des Neuronalen Netzwerkes gestohlen wurden und Yahoo wolle mit der kostenlosen Veröffentlichung verhindern, dass die geklauten Inhalte über Schwarzmarktforen verkauft werden.

Um die Funktion von Caffe und NSFW zu testen, wurde mit Hilfe eines von Yahoo bereitgestellten Dockerfiles ein Dockercontainer erstellt, der das Neuronale Netzwerk sowie das Model beinhaltet. Folgend werden die zuvor in Kapitel 5.1.3 genutzten Testdateien verwendet: zehn Portraits von Schauspielern sowie zehn pornografische Bilder.

---

[15] Jay Mahadeokar, Gerry Pesavento, "Open Sourcing a Deep Learning Solution for ... | Yahoo Engineering", https://yahooeng.tumblr.com/post/151148689421/open-sourcing-a-deep-learning-solution-for, angesehen am 10.Juli 2017

[16] Jay Mahadeokar, Gerry Pesavento, "Open Sourcing a Deep Learning Solution for ... | Yahoo Engineering", https://yahooeng.tumblr.com/post/151148689421/open-sourcing-a-deep-learning-solution-for, angesehen am 10.Juli 2017

Erstere erzielen sehr geringe Werte zwischen 0.000058 und 0.0193. Damit sind sie eindeutig nicht pornografischen Ursprungs. Einzig Dat8 erhält einen NSFW-Score von 0.3444, der deutlich über den anderen Werten liegt, jedoch noch immer weit unterhalb des von Yahoo empfohlenen Grenzwertes liegt.

Datei-name. jpg	**Name Schau-spieler**	**Quelle**	**NSFW-Score**
Dat1	Robert Downey Jr.	https://s-media-cache-ak0.pinimg.com/originals/da/dc/d4/dadcd453713f86372e4297352939976b.jpg	0.000058
Dat2	Kiefer Sutherland	https://image.gala.de/20519312/large1x1-460-460/4243f2a 58b6bfbf09e7f78cc25509967/JN/kiefer-sutherland--10016056-.jpg	0.0008
Dat3	Tom Hiddleston	https://upload.wikimedia.org/wikipedia/commons/3/39/Tom_Hiddleston_in_2013.jpg	0.0001
Dat4	Brad Pitt	http://img.usmagazine.com/article-leads-vertical-300/1250530894_brad_pitt_290x402.jpg	0.000077
Dat5	Sylvester Stallone	https://image.gala.de/20520344/2x3-620-930/e4b46564bdf69a2c9ee0f56d182ee8b7/HT/sylvester-stallone-cm--8555947-.jpg	0.0018
Dat6	Angelina Jolie	http://media.vanityfair.com/photos/57e15c417dd0d7d276c7cb7c/master/h_590,c_limit/angelina-jolie-vf-december-2014-ss02.jpg	0.0193
Dat7	Helen Mirren	http://thatmomentin.com/wp-content/uploads/2016/05/helen-mirren.jpg	0.0128
Dat8	Jenifer Lawrence	http://img.usmagazine.com/article-leads-vertical-300/1298054821_jennifer-lawrence-402.jpg	0.3444
Dat9	Nathalie Dormer	https://upload.wikimedia.org/wikipedia/commons/b/b2/Natalie_Dormer_2014.jpg	0.0030
Dat10	Robin Wright	https://www.welt.de/img/vermischtes/mobile138426852/1622500487-ci102l-w1024/RTX18SIE-jpg.jpg	0.0140

Tabelle 21 Test 1 von Caffe/NSFW mit zehn Schauspielerportraits, Quellen angesehen am 4. Juli 2017, eigene Quelle

Die zehn pornografischen Bilder erhalten hingegen je einen Score zwischen 0.9682 und 0.9999. Sie werden damit eindeutig der Pornografie zugeordnet und liegen weit oberhalb des Grenzwertes. Die Genauigkeit aller getesteten Bilder beträgt damit 100%.

Dateiname.jpg	**Quelle**	**NSFW-Score**
Dat11	http://208.94.232.100/np/thumbs/uf/332607.jpg	0.9999
Dat12	http://208.88.225.212/np/thumbs/rf/329556.jpg	0.9999
Dat13	http://199.80.52.156/np/thumbs/Bf/339848.jpg	0.9682
Dat14	http://208.94.232.100/np/thumbs/207/207012.jpg	0.9874
Dat15	http://208.94.232.100/np/thumbs/Mf/350122.jpg	0.9888
Dat16	http://208.88.225.212/np/thumbs/215/215121.jpg	0.9968
Dat17	http://199.80.52.149/np/thumbs/df/315729.jpg	0.9991
Dat18	http://208.88.225.212/np/thumbs/212/212941.jpg	0.9823
Dat19	http://208.88.225.212/np/thumbs/248/248726.jpg	0.9999
Dat20	http://199.80.52.156/np/thumbs/If/346838.jpg	0.9825

Tabelle 22 Test 2 von Caffe/NSFW mit zehn pornografischen Bildern, Quellen angesehen am 4. Juli 2017 eigene Quelle

Zusätzlich wurden, wie schon im Test von RedLight, je ein Bild in ein Schwarz-Weiß-Format konvertiert und NSFW übergeben. Die Werte von 0.000044 und 0.9974 sind für ihre Kategorie jeweils typisch und beweisen, dass das Neuronale Netzwerk auch Schwarz-Weiß-Bilder erkennen kann. Hier zeigt sich ein weiterer deutlicher Vorteil gegenüber RedLight.

Die Erkennung von kinder- und jugendpornografischen Inhalten kann im Rahmen dieser Arbeit leider nicht getestet werden, da das NSFW-Model nicht auf die Unterscheidung kinder- und jugendpornografischen Inhalte von Erwachsenenpornografie angelernt wurde. Allerdings ist stark davon auszugehen, dass diese inkriminierten Inhalte ebenso wie pornografische Bilder erkannt werden. Somit wäre erneut ein erhebliche Zeitaufwand gespart, müsste der Ermittler nur noch die von NSFW gefilterten Bilder ansehen.

Es ist leider nicht möglich TRAFFIC und Caffe/NSFW direkt gegenüberzustellen. Prinzipiell besitzt die Fraunhofer-Lösung ein hilfreiches und relevantes Feature, dass NSFW nicht kennt. Jedoch funktioniert dieses recht schlecht und erzielt unverwertbare Resultate. Beide sind nicht in der Lage Videodateien zu verarbeiten.

Aufgrund der gemischten Resonanz auf die zuvor erwähnten Lösungen und die dennoch hohe Meinung der Wissenschaft bezüglich Neuronaler Netzwerke, wird die DigiFors GmbH in Zukunft basierend auf Yahoos Caffe oder TensorFlow eine eigene "deep-learning machine" anlernen und testen.

Der testweise Ansatz dabei soll nicht sein (kinder-)pornografische Inhalte zu suchen, da dieser Vorgang sich als schwierig erwiesen hat und hohe Fehler erster und zweiter Art hervorruft. Vielmehr ist gedacht die Maschine nach eindeutig nicht-pornografischen Inhalten zu trainieren und diese ohne Fehler zweiter Art zu filtern. Somit könnte ein Großteil irrelevanter Dateien aussortiert werden.

5.2 Vergleich mit manueller Arbeit eines IT-Forensikers

Alle 54 von der Analysesoftware untersuchten Dateien werden einem menschlichen Ermittler vorgelegt, der seit einigen Jahren an der Aufklärung von Verbrechen im Bereich der Kinder- und Jugendpornografie arbeitet. In unzähligen Fällen mussten dazu sämtliche vorgefundene Foto-, Audio- und Videodateien per Hand durchgesehen und jede einzelne auf inkriminierte Inhalte untersucht werden.

Dies ist eine zeitaufwendige Arbeit, die sowohl Ressourcen zehrt als auch auf Dauer emotional belastend ist. Doch zum jetzigen Zeitpunkt ist dies die State-of-the-Art-Methode, die von Ermittlern auf der ganzen Welt praktiziert wird.

Des Weiteren ist zu erwähnen, dass die Altersschätzung und Geschlechtsbestimmung subjektiv beeinflusst wird. Ein- und dieselbe Datei kann von verschiedenen Ermittlern unterschiedlich bewertet werden, und da die Altersschätzung schon bei geringfügigen Differenzen zu einer Einordnung in eine andere Altersgruppe führt, werden fallrelevante Entscheidungen subjektiv beeinflusst.

Ein eigentlicher Fall von Kinderpornografie kann durch die Überschätzung des Alters des Kinders als Jugendpornografie eingeordnet werden und damit verfahrenstechnisch-entscheidende Änderungen mit sich bringen, wie Strafmaß und Strafdauer. Noch gravierender ist die Einschätzung eines Videos mit Jungendpornografie als legale Erwachsenpornografie, da hierbei im Normalfall keinerlei Straftat besteht.

Die Verwendung eines präzisen und vollautomatischen Tools, das noch dazu völlig objektiv operiert, würde eindeutige Ergebnisse liefern und eventuell bestehende Streitfragen oder Uneinigkeiten der Altersschätzung aus dem Weg räumen.

Für die gerichtsfeste Verwertbarkeit ist jedoch eine Software mit sehr hoher Genauigkeit und geringer Fehlerrate entscheidend.

Die vom Ermittler vorgelegten Ergebnisse werden gleich denen der Analysesoftware ausgewertet und nachfolgend verglichen.

5.2.1 Geschlechtsbestimmung

Die Analyse der Videodateien durch den Ermittler ließ eine Erkennung aller 54 Dateien zu, auch der für die Software problematischen Dat20. Es wurden somit 30 männliche und 24 weibliche Schauspieler analysiert.

Geschlecht	**Erkannt Männlich**	**Erkannt Weiblich**
Männlich (30)	28	2
Weiblich (24)	1	23

Tabelle 23 Anzahl männlicher und weiblicher Teilnehmer und die Verteilung der Geschlechtsbestimmung durch den Ermittler, eigene Quelle

28 der 30 männliches Subjekte wurden richtig erkannt, nur zwei wurden falsch klassifiziert. Die Genauigkeitsrate beträgt 93,3% und liegt damit weit über dem Wert der Software (58,6%). Bei den zwei falsch erkannten Dateien handelt es sich um Dat13 und Dat14, die beiden Szenen mit Jaden Smith, der im sehr jungen Alter längere Haare und Rastazöpfe trug. Dies könnte die Fehleinschätzung des Ermittlers erklären.

Bei den weiblichen Dateien wurden 23 von 24 richtig erkannt, nur eine einzige wurde den männlichen Schauspielern zugeordnet Dies macht eine Genauigkeit von 95,8% aus. Sie liegt geringfügig unterhalb des Erkennungswertes von SHORE (100%).

Insgesamt ist das erkennen des Geschlechts eine für den Menschen recht einfache Aufgabe und wird sowohl bei weiblichen als auch bei männlichen Subjekten mit geringen Fehlerraten absolviert. Es existieren jedoch geringe Fehlerquellen, besonders im Kindesalter.

5.2.2 Altersbestimmung

Geschlecht	Alter genau geschätzt	Alter zu hoch geschätzt	Alter zu gering geschätzt
Männlich(30)	8	10	12
Weiblich(24)	4	18	2

Tabelle 24 Geschlechterabhängige Altersbestimmung unterteilt nach exakter, zu hoher oder zu niedriger Altersschätzung, eigene Quelle

Es wurde bei acht der 30 männlichen Tests das exakte Alter bestimmt, was eine Genauigkeit von 26,7% ausmacht. Zehn Schauspieler wurden älter eingeschätzt als sie es sind, zwölf Weitere hingegen jünger. Dies sind Anteile von 33,3% respektive 40%. Wie bei der Softwareerkennung ist festzustellen, dass ein sehr geringfügiger Teil in Summe jünger geschätzt wird, als älter. Jedoch ist der Anteil an exakt Bestimmten deutlich höher (Vergleich: 3,4%).

Von den 24 Weiblichen Personen wurden vier exakte Alter bestimmt, das macht 16,7% aus. Dieser Wert liegt leicht über dem der Software (12,5%). Es wurden 18 Subjekte älter eingeschätzt als sie es in Wahrheit sind, zwei Weitere wurden jünger eingeschätzt. Die Anteile belaufen sich auf 75% und 8%. Ersteres entspricht dem Wert den auch die Softwarelösung erzielte, letzteres liegt etwas darunter. Damit ist offensichtlich, dass sowohl Ermittler als auch SHORE dazu tendieren weibliche Personen im Alter nach oben zu stufen. Die größte Differenz ist mit acht Jahren erneut im Bereich der weiblichen Jugendlichen zu finden, es handelt sich jedoch um andere Dateien als die, bei denen die Software große Altersdifferenzen aufwies.

Es folgt die Untersuchung der einzelnen Altersgruppen und die jeweilige Berechnung der drei Maße: Precision, Recall und F_1-Maß.

Männl. Kinder	**Erkannt**	**Nicht Erkannt**
Sind (20)	16	4
Sind nicht (10)	0	10

Tabelle 25 Erkennung der männlichen Kinder, eigene Quelle

Genau wie die Gesichtserkennungssoftware werden 16 der 20 männlichen Kinder erkannt, was eine Genauigkeit von 80% ausmacht. Der Fehler der ersten Art beträgt 20%. Ein Fehler der zweiten Art existiert nicht, da die 10 männlichen Jugendlichen und Erwachsenen nicht fälschlicherweise als Kinder klassifiziert werden.

Maß	**Ergebnis**	**Berechnung**
Precision	1	16/(16+0)
Recall	0,8	16/(16+4)
F1-Maß	0,888888889	2*1*0,8/(1+0,8)

Tabelle 26 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der männlichen Kinder, eigene Quelle

Die Präzision berechnet sich aus 16/(16+0) und ergibt genau Eins, die Sensitivität errechnet sich aus 16/(16+4) und ergibt 0,8. Das F_1-Maß wird kalkuliert durch 2*1*0,8/(0,8+1) und ergibt 0,89. Da sich im Vergleich mit SHORE der Recall nicht verändert wurde, aber die Präzision höher ist, erlangt das F_1-Maß ein leicht höheres Ergebnis.

Männl. Jugendliche	**Erkannt**	**Nicht Erkannt**
Sind (4)	4	0
Sind nicht (26)	4	22

Tabelle 27 Erkennung der männlichen Jugendlichen, eigene Quelle

Alle vier männlichen Jugendlichen wurden richtig klassifiziert. hier zeigt sich eine extreme Differenz zur Leistung der Software, die 0% erkannt hat. Zusätzlich wurden vier Kinder fälschlicherweise als Jugendliche erkannt. Dies entspricht 13,6% und damit demselben Wert wie bei der Analyse durch die Software.

Maß	Ergebnis	Berechnung
Precision	0,5	4/(4+4)
Recall	1	4/(4+0)
F1-Maß	0,666666667	2*1*0,5/(1+0,5)

Tabelle 28 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der männlichen Jugendlichen, eigene Quelle

Die Präzision berechnet sich aus 4/(4+4) und ergibt 0,5, die Sensitivität errechnet sich aus 4/(4+0) und ergibt 1. Das F_1-Maß wird kalkuliert durch 2*1*0,5/(1+0,5) und ergibt 0,66. Dieses Ergebnis ist deutlich besser als das der Software (0) und wird durch die vier richtig klassifizierten männlichen Jugendlichen Hervorgerufen.

Männl. Erwachsene	Erkannt	Nicht Erkannt
Sind (6)	6	0
Sind Nicht (24)	0	24

Tabelle 29 Erkennung der männlichen Erwachsenen, eigene Quelle

Alle sechs männlichen Erwachsenen wurden richtig klassifiziert, damit liegt die Trefferrate bei 100%. Kein männliches Kind oder Jugendlicher wurde der Gruppe der Erwachsenen zugeordnet, daher sind sowohl der Fehler erster Art als auch der Fehler zweiter Art non-existent.

Maß	Ergebnis	Berechnung
Precision	1	6/(6+0)
Recall	1	6/(6+0)
F1-Maß	1	2*1*1/(1+1)

Tabelle 30 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der männlichen Erwachsenen, eigene Quelle

Die Präzision berechnet sich aus 6/(6+0) und ergibt 1, die Sensitivität errechnet sich aus 6/(6+0) und ergibt 1. Das F_1-Maß wird kalkuliert durch 2*1*1/(1+1) und ergibt ebenso 1. Dieses Ergebnis ist eindeutig nicht verbesserbar.

Allerdings ist die geringe Menge an untersuchten Daten zu bedenken und die Tatsache, dass eine zu hohe Einstufung des Alters in der Erwachsenengruppe nicht möglich ist.

Weibl. Kinder	**Erkannt**	**Nicht Erkannt**
Sind (6)	6	0
Sind nicht (18)	2	16

Tabelle 31 Erkennung der weiblichen Kinder, eigene Quelle

Mit einer Genauigkeit von 100% wurden alle sechs weiblichen Kinder der richtigen Altersgruppe zugeordnet. Des Weiteren wurden zwei Kinder als Jugendliche eingeschätzt (unkritischer Fehler). Von den weiblichen Jugendlichen und Erwachsenen macht das einen Anteil von 11,1% aus. Dieser Wert entspricht dem, der Softwareanalyse.

Maß	**Ergebnis**	**Berechnung**
Precision	0,75	6/(6+2)
Recall	1	6/(6+0)
F1-Maß	0,857142857	2*1*0,75/(1+0,75)

Tabelle 32 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der weiblichen Kinder, eigene Quelle

Die Präzision berechnet sich aus 6/(6+2)und ergibt 0,75, die Sensitivität errechnet sich aus 6/(6+0) und ergibt 1. Das F_1-Maß wird kalkuliert durch 2*1*0,75/(1+0,75) und ergibt 0,86. Es werden alle gesuchten Personen richtig erkannt, jedoch auch Ergebnisse genannt, die nicht dazu gehören. Es existiert demnach kein Fehler erster Art, aber ein Fehler zweiter Art. Der Wert des F_1-Maßes beläuft sich in ähnlichem Bereich wie schon zuvor bei den männlichen Kindern und ist identisch mit dem Erkennungswert der weiblichen Kinder durch SHORE.

Weibl. Jugendliche	**Erkannt**	**Nicht Erkannt**
Sind (14)	6	8
Sind nicht (10)	0	10

Tabelle 33 Erkennung der weiblichen Jugendlichen, eigene Quelle

Bei den Tests mit weiblichen Jugendlichen sind die größten Unterschiede zwischen menschlichem Ermittler und Gesichtserkennungssoftware aufgetreten. Der Ermittler erkennt sechs der 14 Subjekte richtig und erzielt damit eine Genauigkeit von 42,8%.

Der Vergleichswert der Software liegt lediglich bei 7,1%. Kein weibliches Kind und keine Erwachsene wurden als Jugendliche eingestuft.

Maß	**Ergebnis**	**Berechnung**
Precision	1	6/(6+0)
Recall	0,428571429	6/(6+8)
F1-Maß	0,6	2*1*0,43/(1+0,43)

Tabelle 34 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der weiblichen Jugendlichen, eigene Quelle

Die Präzision berechnet sich aus 6/(6+0) und ergibt 1, die Sensitivität errechnet sich aus 6/(6+8) und ergibt 0,43. Das F_1-Maß wird kalkuliert durch 2*1*0,43/(1+0,43) und ergibt 0,6.

Dieser Wert liegt weit unter den Resultaten, die sowohl Mensch als auch Maschine zum Beispiel bei männlichen und weiblichen Kindern erreichen. Er liegt jedoch weit über dem Ergebnis der Softwareerkennung (0,13).

Es ist erneut offensichtlich, dass die Gruppe der weiblichen Jugendlichen einen problembelasteten Schwerpunkt der Altersschätzung ausmacht. Der menschliche Ermittler kann deutlich besser Resultate erzielen als die Gesichtserkennungssoftware, tendiert jedoch trotzdem dazu einen Großteil der weiblichen Jugendlichen nicht zu erkennen, sondern älter einzustufen. Hier entsteht somit ein unkritischer Fehler.

Erwachsene	**Erkannt**	**Nicht Erkannt**
Sind (4)	4	0
Sind Nicht (20)	6	14

Tabelle 35 Erkennung der weiblichen Erwachsenen, eigene Quelle

Alle vier weiblichen Erwachsenen wurden richtig klassifiziert, zusätzlich wurden sechs der 20 weiblichen Kinder und Jugendlichen durch eine zu hohe Einschätzung ihres Alters als Erwachsen angesehen. Dies macht einen Anteil von genau 30%.

Maß	Ergebnis	Berechnung
Precision	0,4	4/(4+6)
Recall	1	4/(4+0)
F1-Maß	0,571428571	2*1*0,4/(1+0,4)

Tabelle 36 Berechnung und Ergebnis von Precision, Recall und F_1-Maß der weiblichen Erwachsenen, eigene Quelle

Die Präzision berechnet sich aus 4/(4+6) und ergibt 0,4, die Sensitivität errechnet sich aus 4/(4+0) und ergibt 1. Das F_1-Maß wird kalkuliert durch 2*1*0,4/(1+0,4)) und ergibt 0,57. Dieser Wert ist besser als das erzielte Ergebnis von SHORE (0,42), liegt jedoch weiterhin unter den angezielten Idealwerten. Auch hier sind die Auswirkungen der mehrfachen Überschätzung des Alters der weiblichen Subjekte erkennbar, da sie die Präzision senken.

5.2.3 Gesamtbewertung

In vier der sechs Tests, nämlich bei männlichen Kindern und Jugendlichen sowie bei weiblichen Jugendlichen und Erwachsenen, erzielt der Ermittler höhere Ergebnisse. Die Differenzen der F_1-Maße belaufen sich auf Werte zwischen 0,06 und 0,67.

Bei männlichen Erwachsenen und weiblichen Kindern erzielen SHORE und die manuelle Analyse des Ermittlers die exakt gleichen Ergebnisse.

Die insgesamt beste Erkennung beider Varianten wurde bei männlichen Erwachsenen mit einem F_1-Maß von Eins erzielt. Die niedrigsten Werte liefert SHORE bei männlichen gefolgt von weiblichen Jugendlichen. Größte Probleme beim Ermittler hingegen treten bei den weiblichen Erwachsenen und ebenso den weiblichen Jugendlichen auf.

Der Ermittler weißt eine durchschnittliche Schätzungsdifferenz zum wahren Alter der Personen von 1,83 Jahren auf und verschätzt sich maximal um acht Jahre. Die Software verschätzt sich um durchschnittlich 1,94 Jahre und gibt Extremabweichungen von maximal zwölf Jahren aus.

Die zeitliche Differenz zwischen den Analysen von Software und Ermittler lässt sich nur schwer an etwas festmachen, da beim Betrachten der GUI mehrere Werte über die Dauer des Videos angezeigt werden und der Benutzer einen Geeigneten heraussuchen muss. Zu welchem Zeitpunkt dieser erscheint, wie oft der Benutzer das Video durchlaufen lässt und welchen Wert der Benutzer individuell für geeignet hält, hängt jeweils vom Benutzer und dem einzelnen Video ab.

Weder der Programmbenutzer noch der Ermittler, der manuell die Videos sichtet, muss das ganze Video anschauen, da sie schon nach wenigen Sekunden ein geeignetes Alter und Geschlecht angezeigt bekommen oder selbst einschätzen können.

Im Rahmen des Tests hat die Analyse der Videodateien mit SHORE circa 30 Minuten gedauert. Der Ermittler benötigte rund 30 Minuten. Damit scheinen beide ungefähr gleichlange zu benötigen.

Erkannt von	Erkannt als	Kinder(26)	Jugendliche(17/18)	Erwachsene(10)
SHORE	Kind	22	5	0
	Jugendliche(r)	4	1	0
	Erwachsene(r)	0	11	10
Ermittler	Kind	22	2	0
	Jugendliche(r)	4	10	0
	Erwachsene(r)	0	6	10
		Altersgruppe richtig geschätzt	Altersgruppe zu hoch geschätzt	Altersgruppe zu gering geschätzt

Tabelle 37 Übersicht Alterserkennung durch SHORE und Ermittler, eigene Quelle

Jedoch wird beim Verwenden von SHORE der zusätzliche Vorbereitungsschritt des Segmentieren gespart, der Videodateien in Einzelbilder unterteilt und den der Ermittler vor der manuellen Alters- und Geschlechtsbestimmung durchführen muss. Dann kann jedoch jedes Einzelne ausreichen um relativ genaue Angaben bezüglich Alter und Geschlecht zu tätigen. Diese zusätzlich benötigte Zeit muss hinzugerechnet werden.

Tabelle 37 ermöglicht einen zusammenfassenden Überblick über die Einteilung der Personen nach Altersgruppen durch die Erkennungssoftware SHORE und den menschlichen Ermittler. Beide erzeugen die häufigsten Fehler (sowohl kritisch als auch unkritisch) bei Personen im Alter von 14 bis 17 Jahren.

Weder Software noch Ermittler haben ein Kind den Erwachsenen oder andersherum zugeordnet. Die Altersdifferenzen zwischen wahrem und geschätztem Alter beliefen sich nie auf zwei Altersstufen.

5.3 Ausblick

Weitere umfangreichere Tests mit dem Programm werden benötigt um nach Möglichkeit die erhaltenen Ergebnisse zu bestätigen und gegebenenfalls anzupassen oder zu entkräften. Hierzu wird eine Erweiterung des Testdatensatzes empfohlen, die Dateien anderer Herkunft und Qualität beinhaltet, als auch eine insgesamt größere Masse an Daten stellt.

Es ist außerdem anzumerken, dass vor allem die Länge der Clips variiert werden sollte. Um die Erkennung durch das Analyseprogramm zu erhöhen, wurden die Dateien mit einer Länge von ungefähr 30 Sekunden ausgewählt. Resultate wurden jedoch schon nach Bruchteilen einer Sekunde angegeben.

Der im Vergleich hinzugezogene Ermittler benötigte nur wenige Sekunden für seine Geschlechtsbestimmung und Altersschätzung, was dazu führt, dass beide Ungefähr dieselbe Dauer aufweisen. Fraglich ist, ob dieselben Clips durch verminderte Länge schneller, aber ebenso präzise, vom Programm erkannt werden. Dies könnte einen deutlich zeiteffizienteren Untersuchungsdurchlauf bedeuten.

Andererseits würden deutlich längere Videos einen Mehraufwand bereiten, da der Benutzer hier mehr Material zur Verfügung hat und prüfen muss. Hier ist fraglich ob schon zu Beginn getroffene Analysewerte auf die Dauer des Clips aktuell bleiben oder aber verändert werden, sowie weitere relevante Personen hinzukommen könnten.

Ein prinzipieller Schritt zur weiteren Automatisierung des Vorgehens wäre die Entwicklung eines Tools, dass anders als die GUI das Video nicht während der Analyse anzeigt, sondern lediglich eine .csv-Datei mit den Resultaten erstellt.

Andererseits scheint auch die Entwicklung eines eigenen, selbst antrainierten Neuronalen Netzwerks sinnvoll. Hierzu ist jedoch ein großer Aufwand nötig, und die Resultate können sich als wenig nützlich herausstellen.

5.4 Fazit

Die Gesichtserkennungssoftware SHORE des Fraunhofer Instituts für Integrierte Schaltungen hat sich als das Tool herausgestellt, das für den Versuch dieser Arbeit Kinder und Jugendliche mittels automatisierter Gesichtserkennung in Videos zu analysieren, am besten geeignet ist. Die zugrundeliegenden Algorithmen erkennen Gesichter schnell, umfassend und auch bei mittelmäßiger Videoqualität. Die Bestimmung des Geschlechts und die Schätzung des Alters wurden jedoch nur mangelhaft durchgeführt.

Für die Bestandskraft vor Gericht und die automatische Analyse von Videodateien treten zu viele Fehler bei der Zuordnung der Geschlechter und Altersgruppen auf. Außerdem ist zu bedenken, dass überhaupt erst Gesichter im Material erkennbar sein müssen. Viele Handlungen im Bereich der Kinder- und Jugendpornografie beinhalten jedoch keine Gesichter oder Köpfe, sondern nur Szenen mit menschlichen Unterkörpern oder Rückansichten von Personen, bei denen das Gesicht nicht erkennbar ist. Daher kann die Software prinzipiell nur bei einem Bruchteil der relevanten Daten hilfreich sein.

Den Resultaten dieser Arbeit zufolge, erscheint die Verwendung von Gesichtserkennungssoftwares (noch) nicht präzise und fehlerfrei genug, um in diesem spezifischen Umfeld eingesetzt werden zu können. Wie die weiteren in Kapitel 5.1 vorgestellten Programme dient dieser Ansatz lediglich zur teilweisen Unterstützung, nicht jedoch zur vollautomatischen Analyse. Die manuelle Auswertung ist momentan unabdingbar.

Literaturverzeichnis

1. Brett Shavers, Eric Zimmerman: X-Ways Forensics Practitioner's Guide, Elsevier Inc., 2014

2. Stan Z. Li, Anil K. Jain: Handbook of Face Recognition, Second Edition, Springer-Verlag, 2011

3. Rein-Lien Hsu: Face Detection and Modeling for Recognition, PhD Thesis, Michigan State University, 2002

4. M. Turk, Alex Pentland: Eigenfaces for Recognition, Journal of Cognitive Neuroscience 3 (1991), S 71-86

5. Barak Moghaddam, Alex Pentland: Bayesian Face Recognition, Pattern Recognition, Vol. 33, Nr. 11, S. 1771-1782, 2000

6. Wenyi Zhao et al.: Discriminant Analysis of Principal Components for Face Recognition, Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition, 1998

7. Harry Wechsler: Reliable Face Recognition Methods - System Design, Implementation and Evaluation, Springer-Verlag, 2007

8. Ivan Huerta et al: A Deep Analysis on Age Estimation, Pattern Recognition Letters, Elsevier, 2015

9. Dat Nien Nguyen et al.: Comparative Study of Human Age Estimation with or without Preclassification of Gender and Facial Expression, The Scientific World Journal Volume 2014, 2014

10. Yuyu Liang et al.: A Hierarchical Framework for Facial Age Estimation, Mathematical Problems in Engineering Volume 2014, 2014

11. Andreas Lantis et al.: Comparing different classifiers for automatic age estimation, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) (Volume 34, Issue 1, Feb 2004), 2004

12. G.J. Edwards, C.J. Taylor, T.F. Cootes: Interpreting Face Images using Active Appearance Models, Procs of the 5th European Conference on Computer Vision, S. 581-595, 1998

13. T.F. Cootes, G.J. Edwards, C.J. Taylor: Active Appearance Models, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23, Nr.6, Juni 2001,S. 681-685, 2001

14. T.F. Cootes, C.J. Taylor, D.H. Cooper and J. Graham :Active Shape Models - Their Training And Application, Computer Vision Graphics and Image Understanding, 61, Nr 1, S. 38-59, 1995.

15. Carles Fernández et al.: A Comparative Evaluation of Regression Learning Algorithms for Facial Age Estimation, Face and Facial Expression Recognition from Real World Videos. Lecture Notes in Computer Science, vol 8912, S 133-144, 2014

16. Guodong Guo, Guowang Mu: Joint Estimation of Age, Gender and Ethnicity: CCA vs. PLS, 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2013

17. Geng et al.: Learning from Facial Aging Patterns for Automatic Face Estimation, ACM Conf. on Multimedia, S. 307-316, 2006

18. Wenming Zheng et al.: Facial Expression Recognition Using Kernel Canonical Correlation Analysis (KCCA), IEEE Transactions on Neural Networks (Volume 17, Issue 1, S. 233-238), 2006

19. David R. Hardoon et al.: Canonical correlation analysis; An overview with application to learning methods, Neural Computation Volume 16 Issue 12, Dez. 2004, S, 2639-2664, 2003

20. Chengjun Liu, Harry Wechsler: Independent Component Analysis of Gabor Features for Face Recognition, IEEE Trans. Neural Networks, Vol. 14, Nr. 4, S. 919-928, 2003

21. Yong Gao et al.: Weighted Gabor Features in Unitary Space for Face Recognition, 7th International Conference on Automatic Face and Gesture Recognition, 2006

22. Ning Sun et al.: Gender Classification Based on Boosting Local Binary Pattern, Advances in Neural Networks - ISNN 2006,Third International Symposium on Neural Networks, Chengdu, China, 2006

23. B.A. Golomb, D.T. Lawrence, T.J. Sejnowski: SexNet: A Neural Network Identifies Sex from Human Faces, Advances in Neural Information Processing Systems 3, Morgan Kaufmann, S. 572-577, 1991

24. Guangcheng Zhang et al.: Bosting Local Binary Pattern (LBP)-Based Face Recognition, Advances in Biometric Person Authentication, S. 179-186, 2004

25. Yoav Freund, Robert E. Schapire: A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, Volume 55, Issue 1, S. 119-139, 1997

26. George Azzopardi, Antonio Greco, Mario Vento: Gender recognition from face images with trainable COSFIRE filters, 13th IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS), 2016

27. Eran Eidinger, Roee Enbar, Tal Hassner: Age and Gender Estimatin of Unfiltered Faces, IEEE Transactions on Information Forensics and Security, Volume 9, Nr. 12, 2014

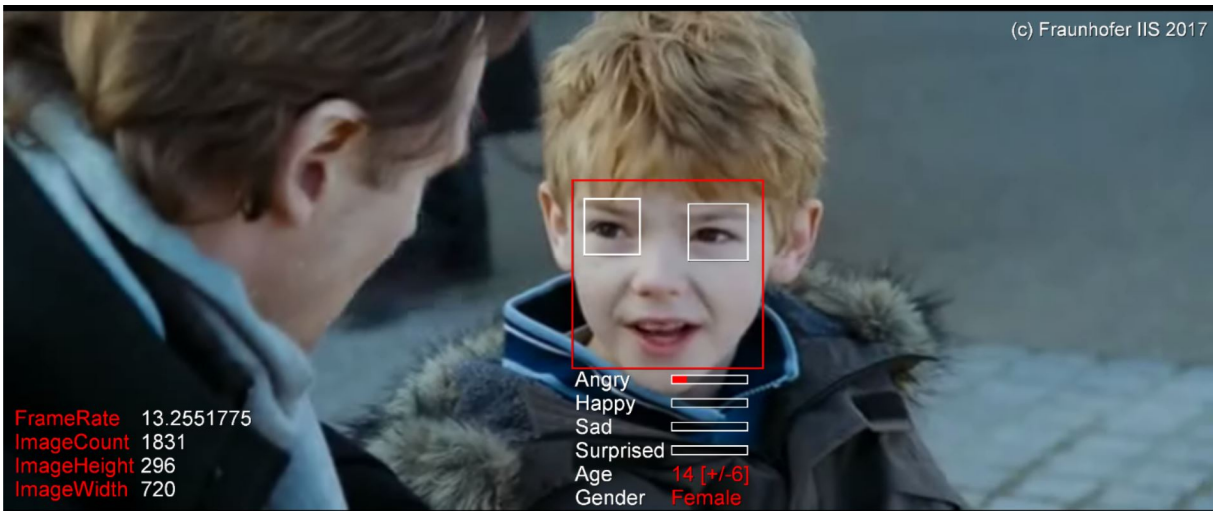
28. Vahid Karimi, Ahkan Tashk: Age and Gender Estimation by Using Hybrid Facial Features, 20th Telecommunications forum TELFOR, 2012

29. Kylie Tanner: Modelling Automated Detection of Children in Images, Master Thesis, University of Rhode Island, 2011

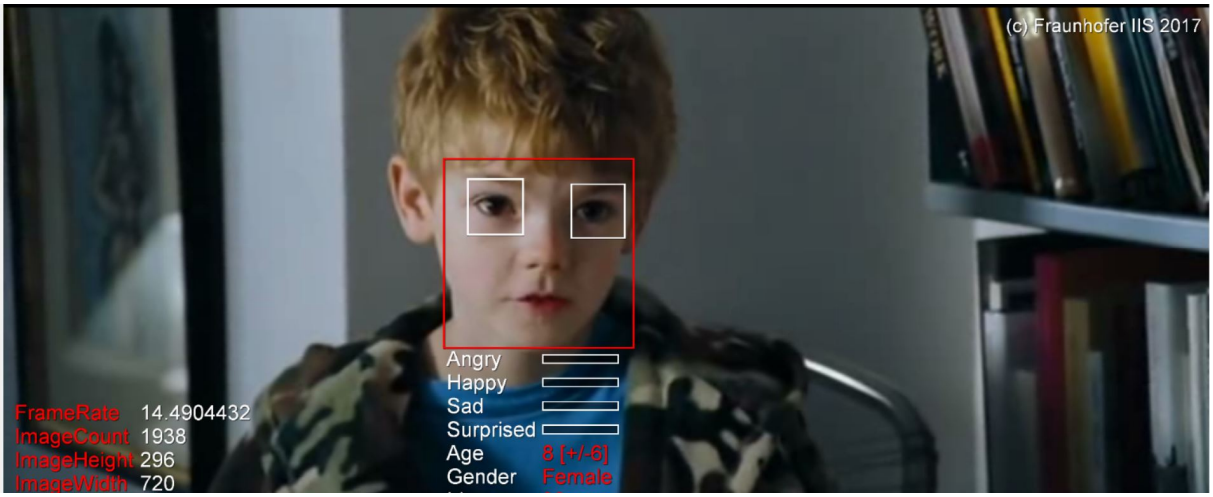
30. Robert J. O'Leary et al. (NIJ/ECTCoE): RedLight (Beta) Software Version 0.1.0.0 Evaluation Report, Januar 2011

31. Timothy H. Ball: Rapid Determination of Age Classification, Dissertation, University of Rhode Island 2011

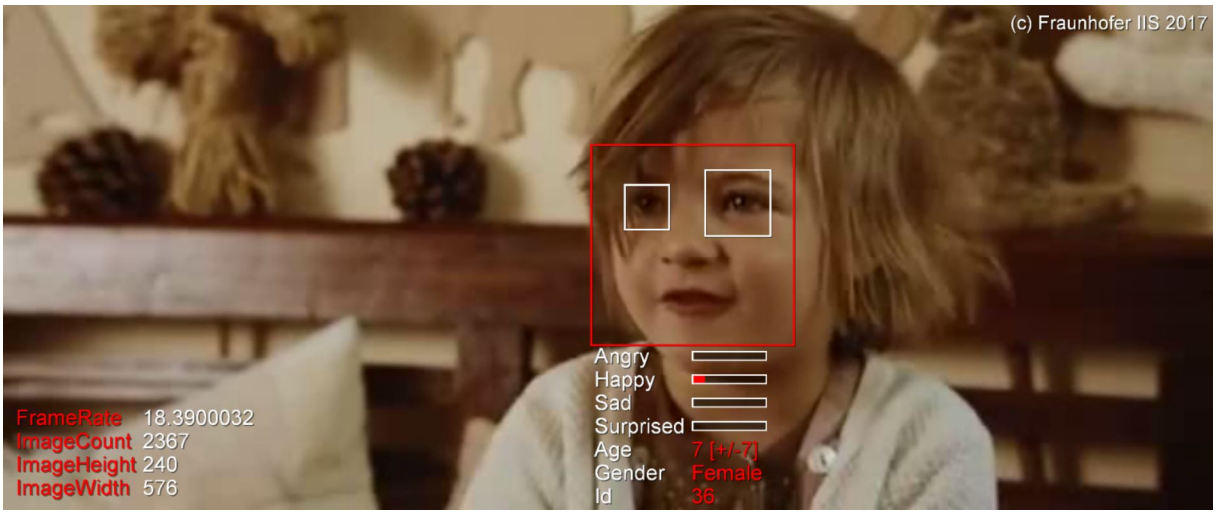
Anlagen



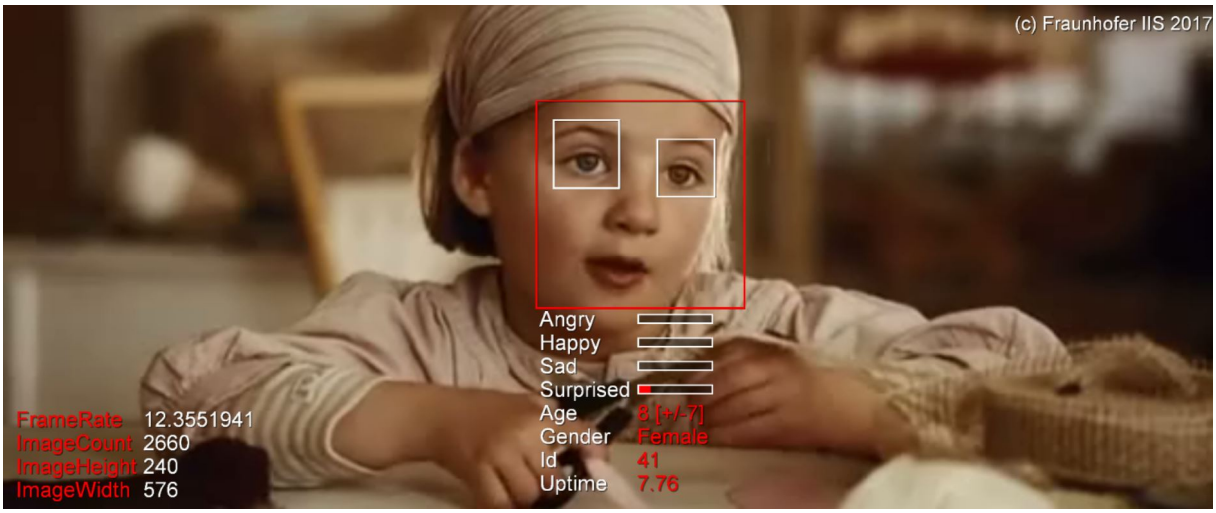
Anhang 1 SHORE-Analyse Dat1



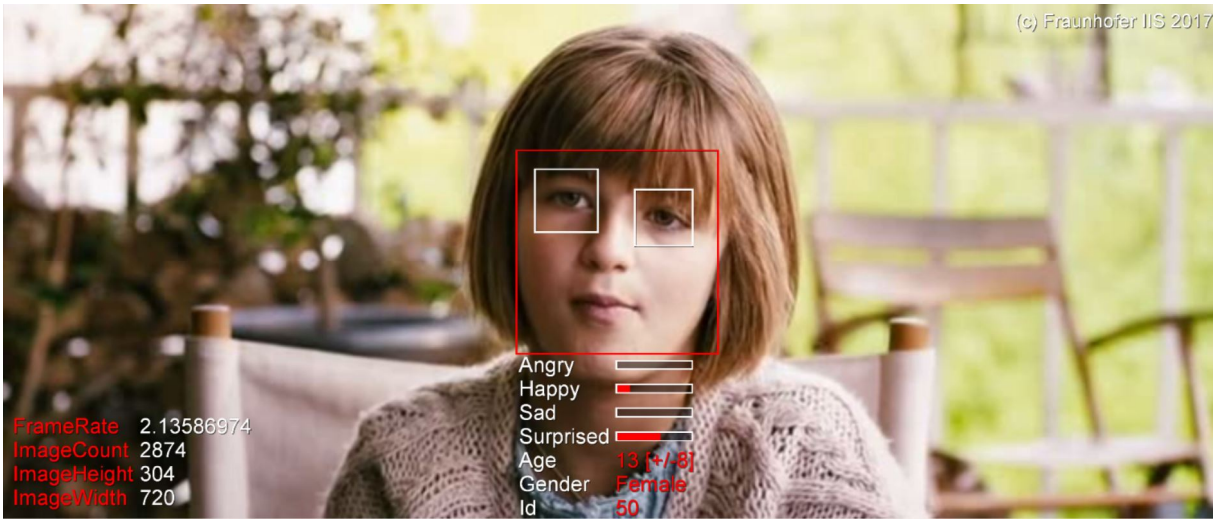
Anhang 2 SHORE-Analyse Dat2



Anhang 3 SHORE-Analyse Dat3

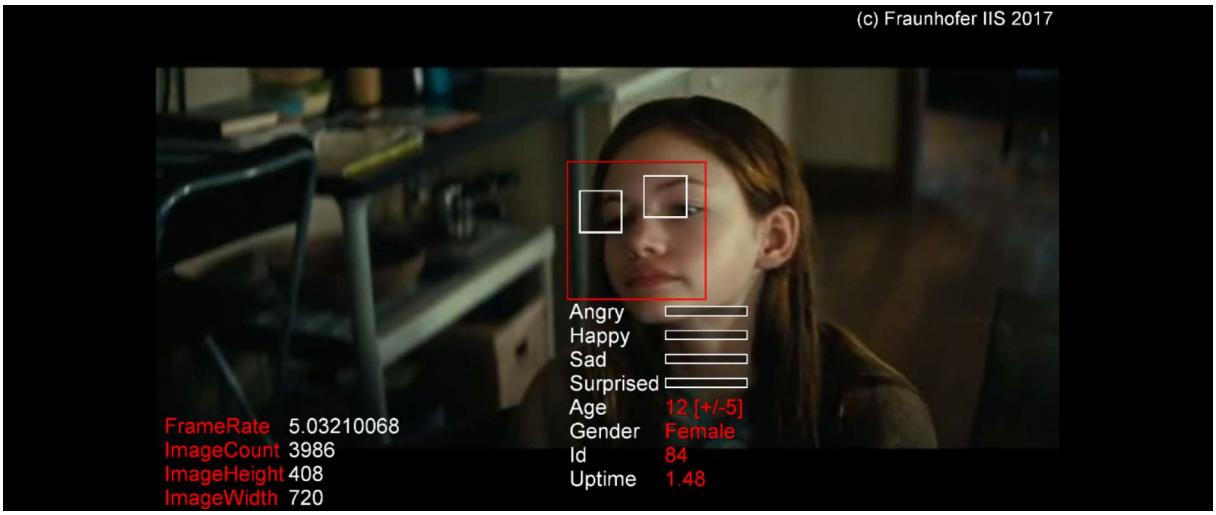


Anhang 4 SHORE-Analyse Dat4

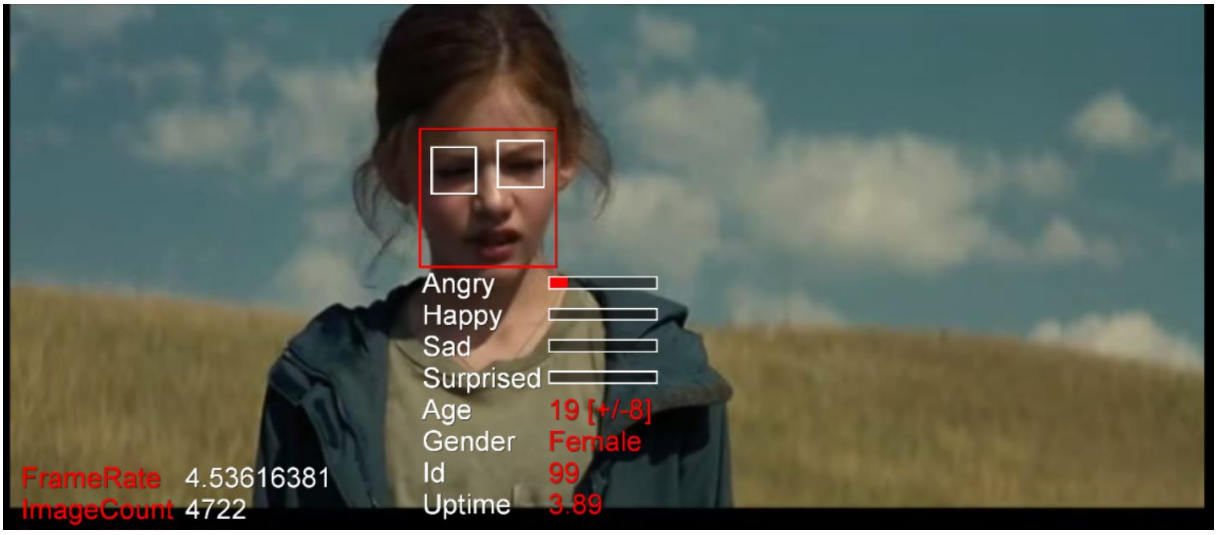


Anhang 5 SHORE-Analyse Dat5

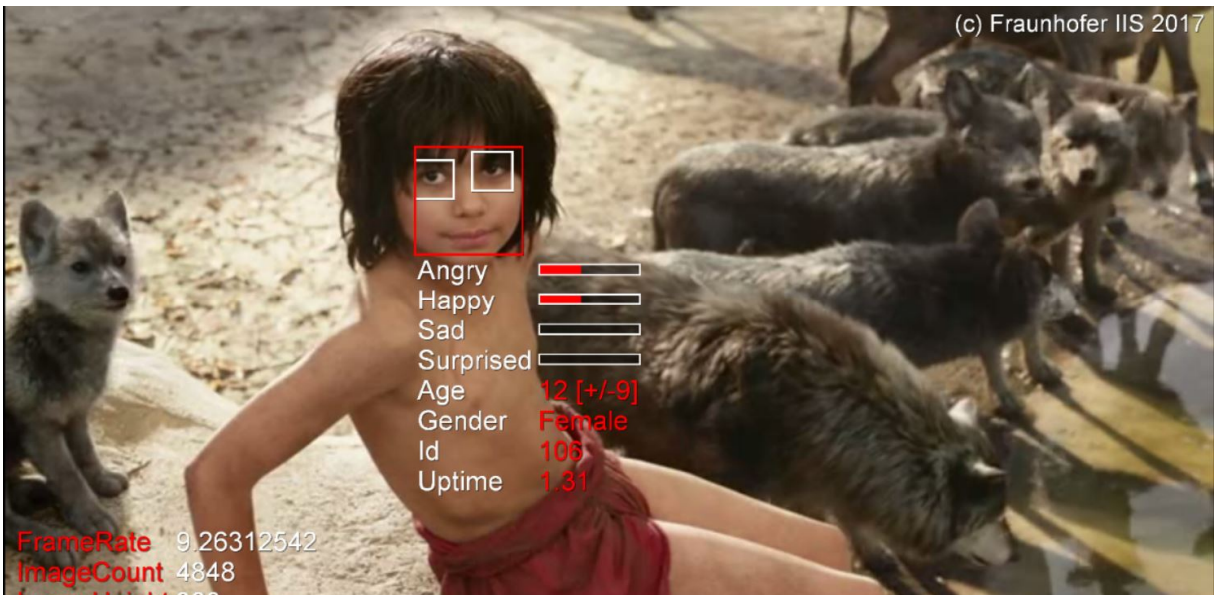
Anhang 6 SHORE-Analyse Dat6



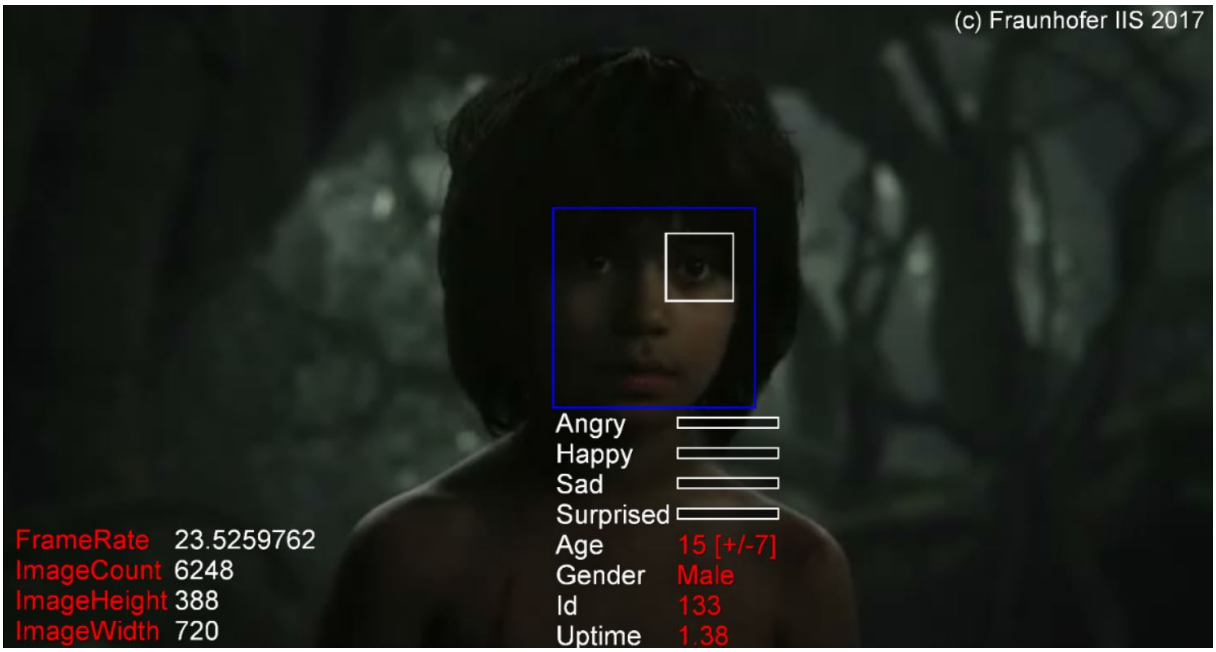
Anhang 7 SHORE-Analyse Dat7



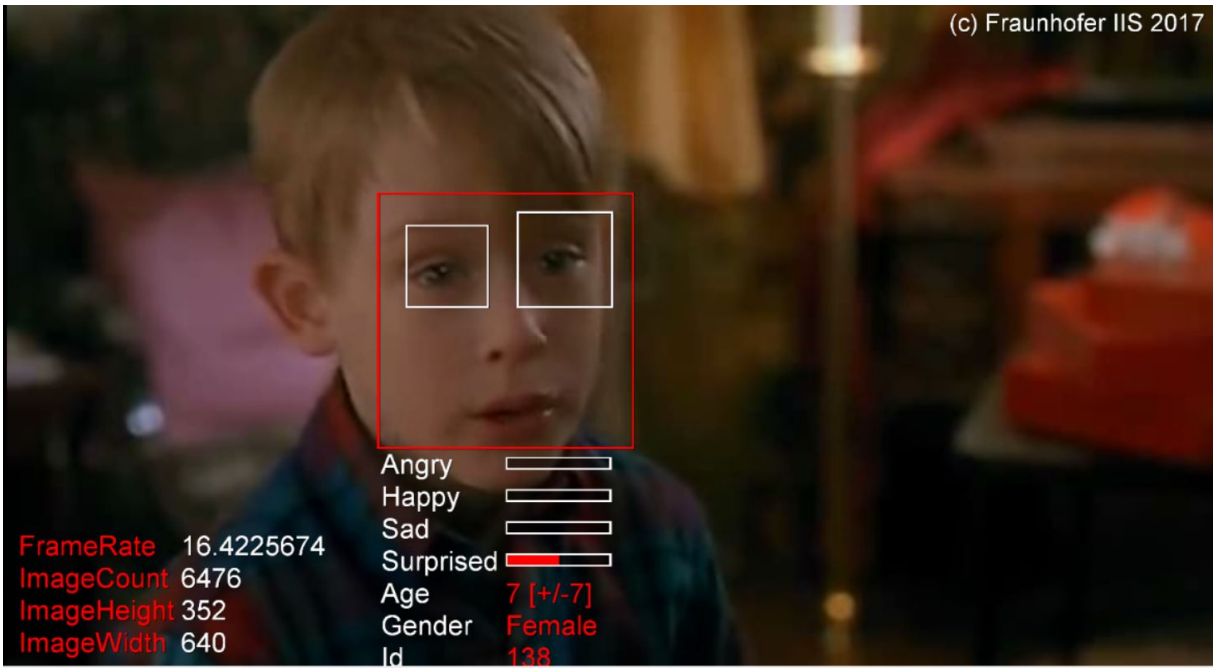
Anhang 8 SHORE-Analyse Dat8



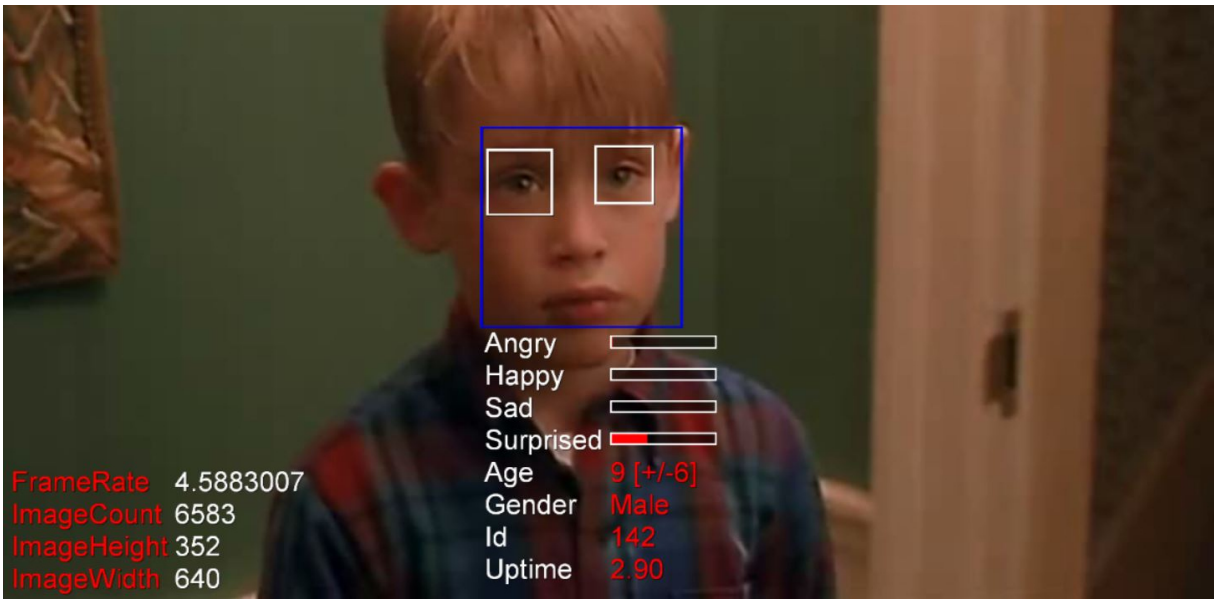
Anhang 9 SHORE-Analyse Dat9



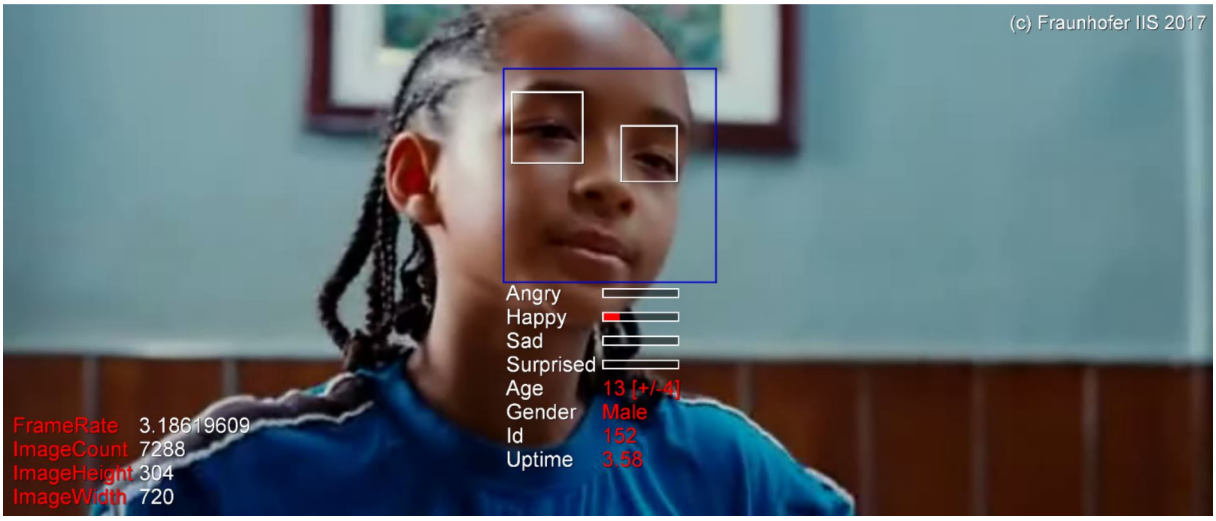
Anhang 10 SHORE-Analyse Dat10



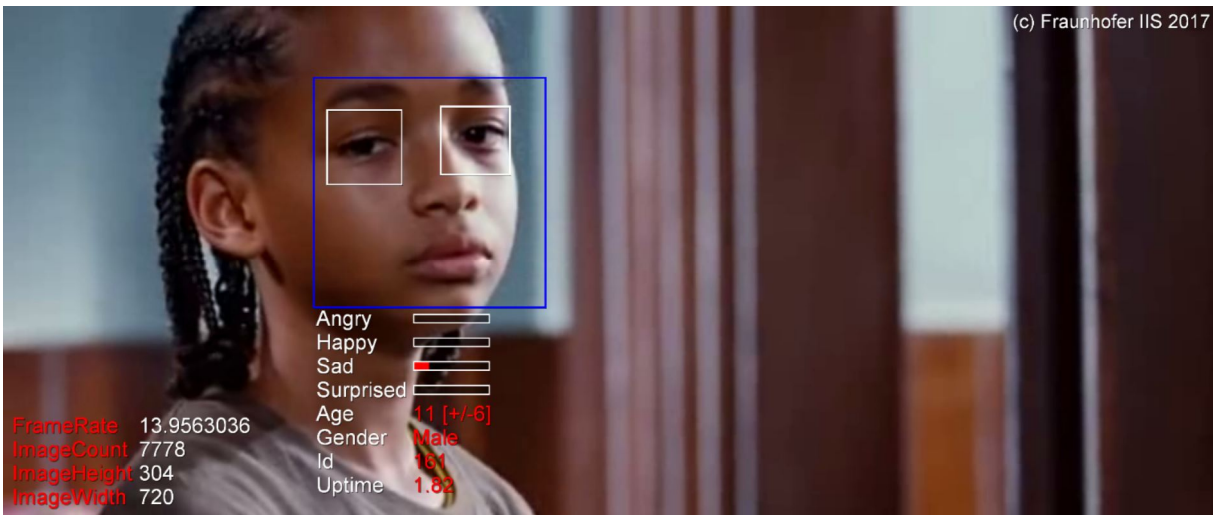
Anhang 11 SHORE-Analyse Dat11



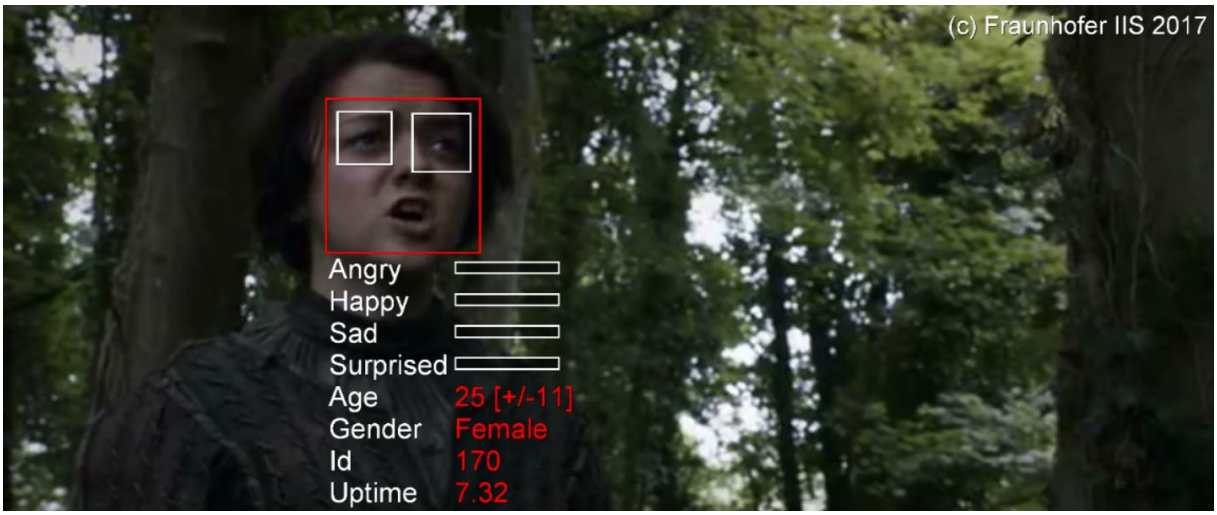
Anhang 12 SHORE-Analyse Dat12



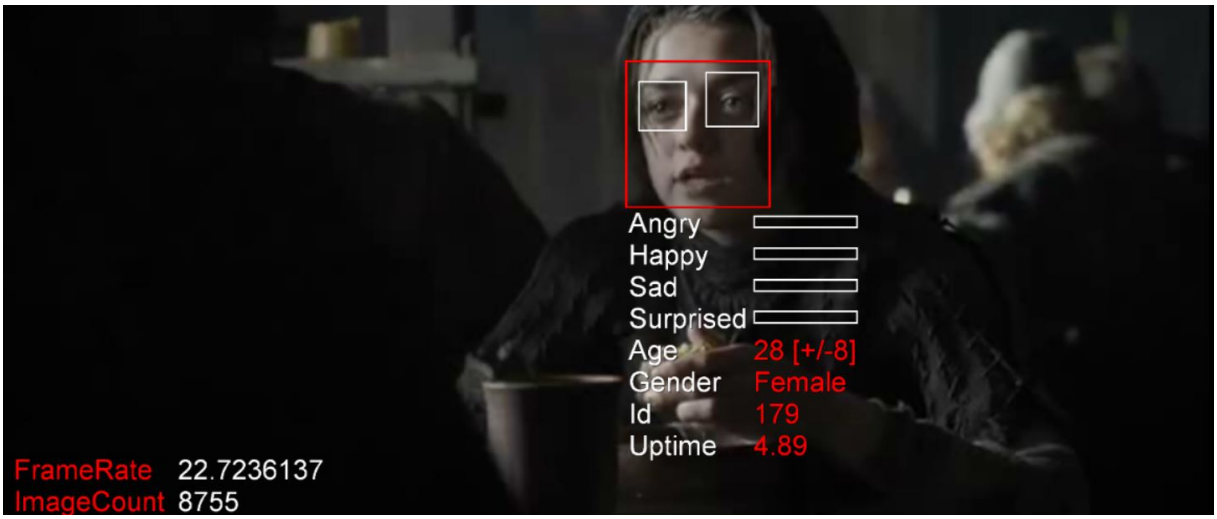
Anhang 13 SHORE-Analyse Dat13



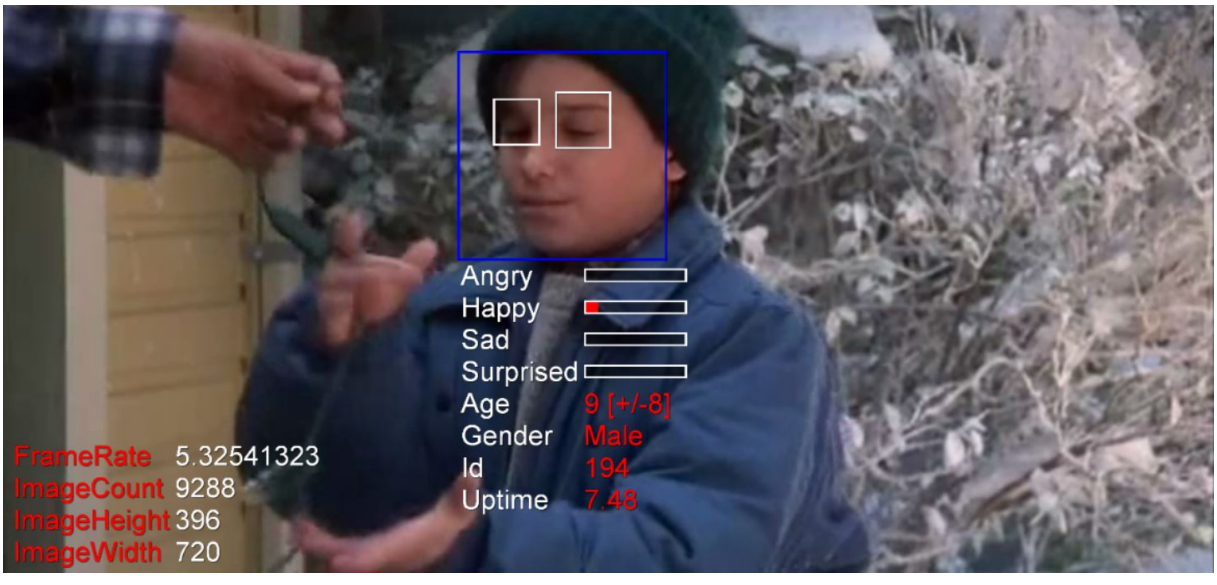
Anhang 14 SHORE-Analyse Dat14



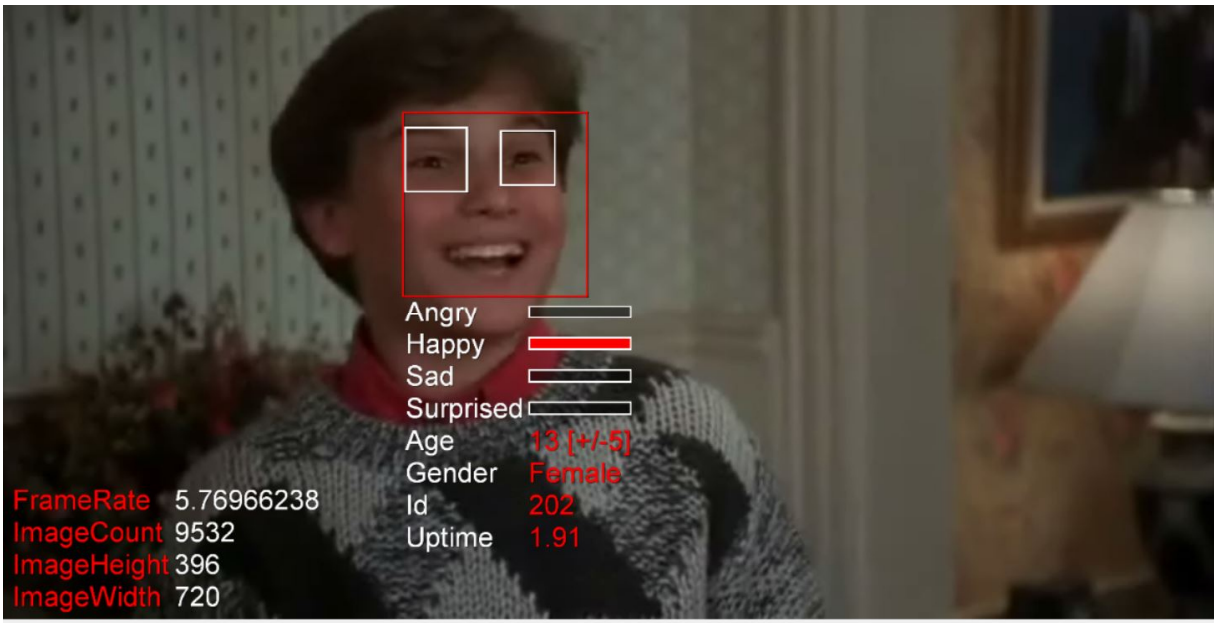
Anhang 15 SHORE-Analyse Dat15



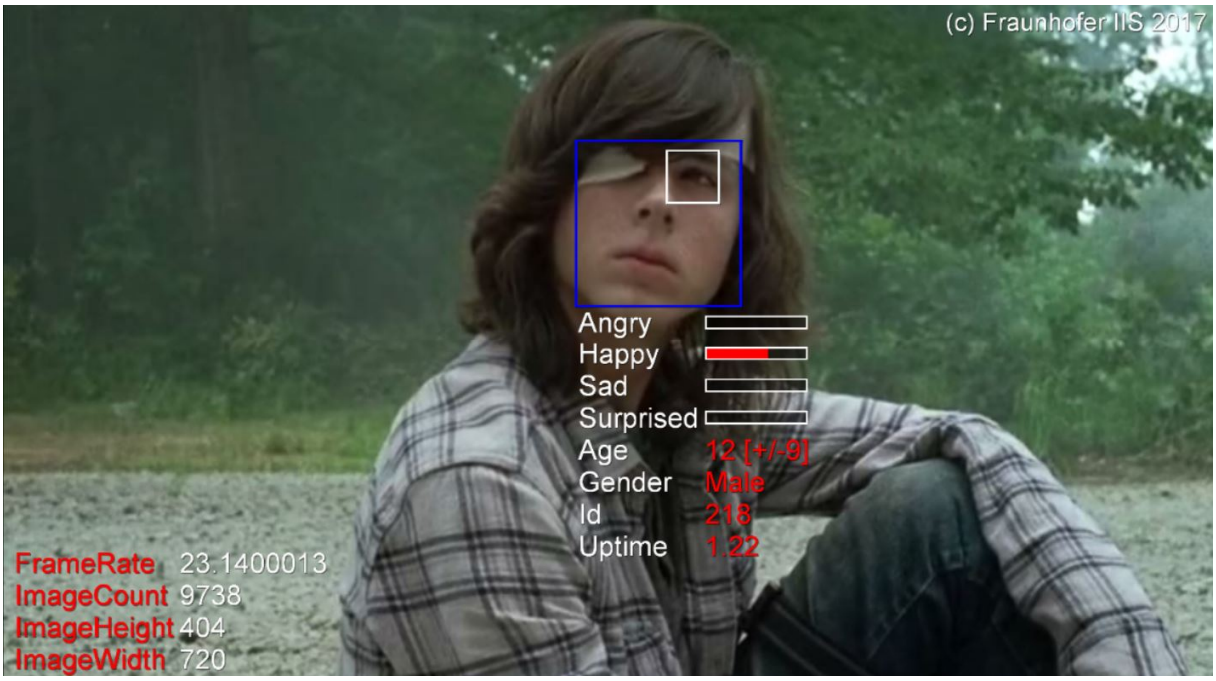
Anhang 16 SHORE-Analyse Dat16



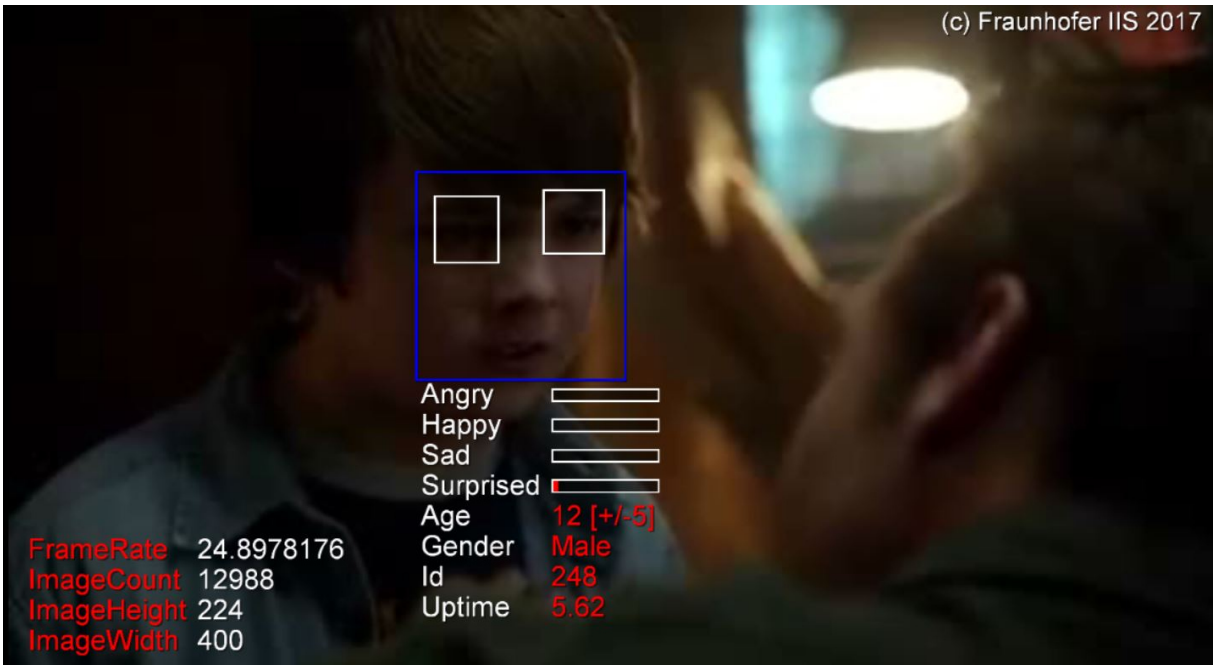
Anhang 17 SHORE-Analyse Dat17



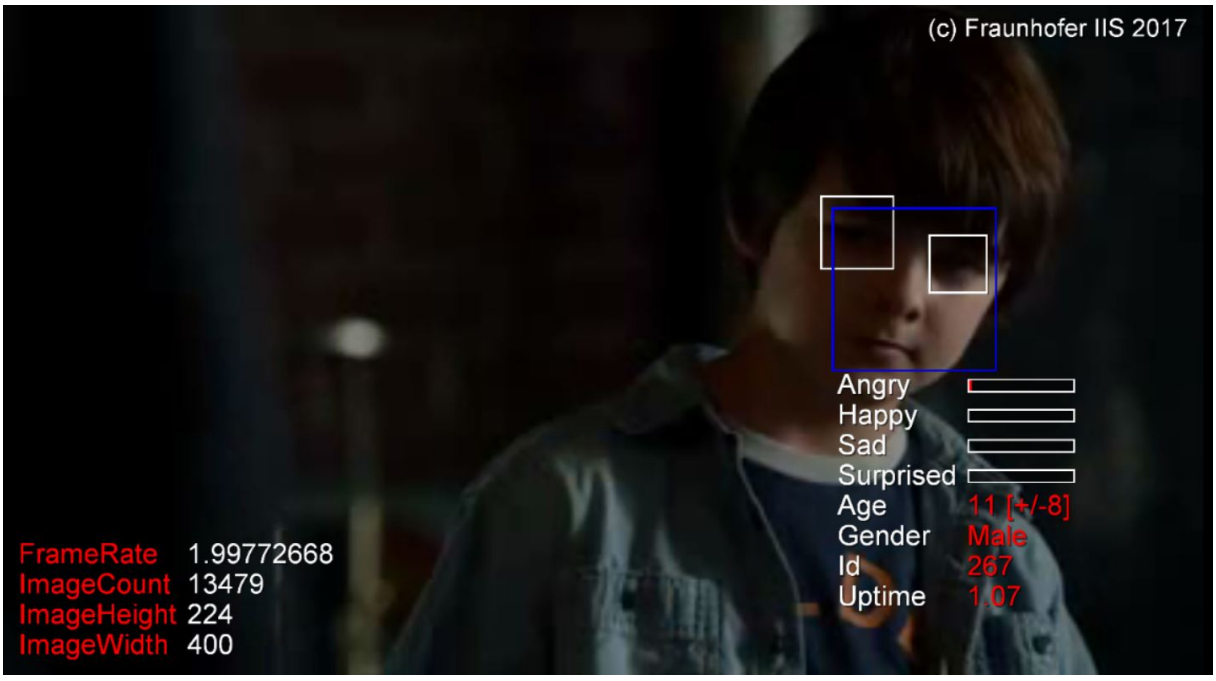
Anhang 18 SHORE-Analyse Dat18



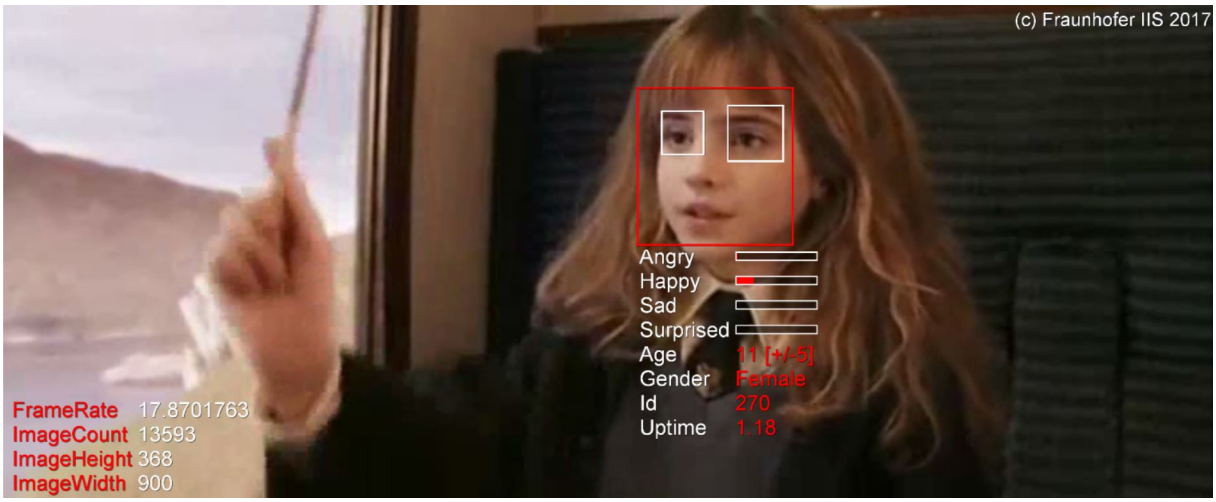
Anhang 19 SHORE-Analyse Dat19



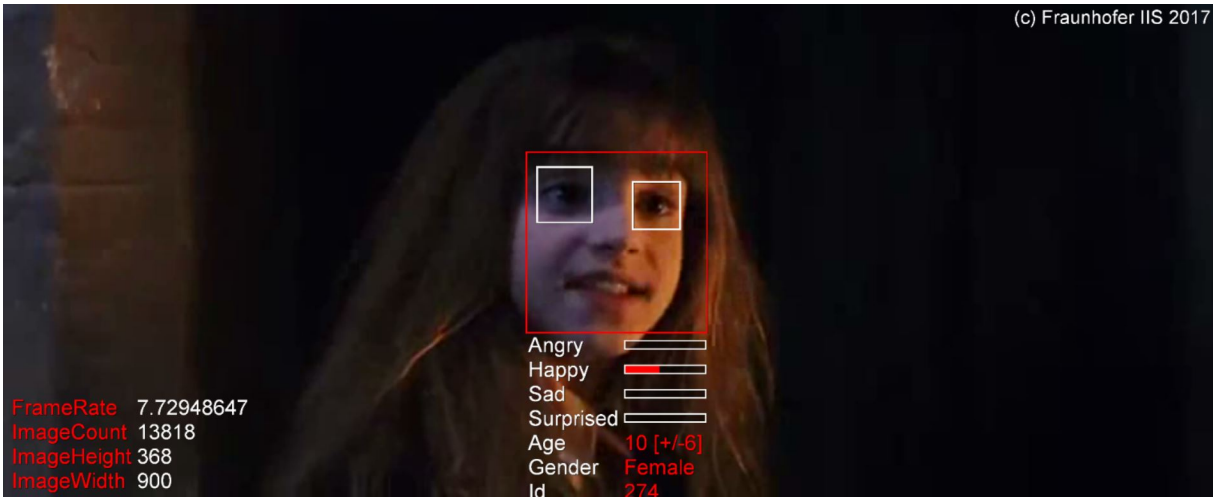
Anhang 20 SHORE-Analyse Dat21



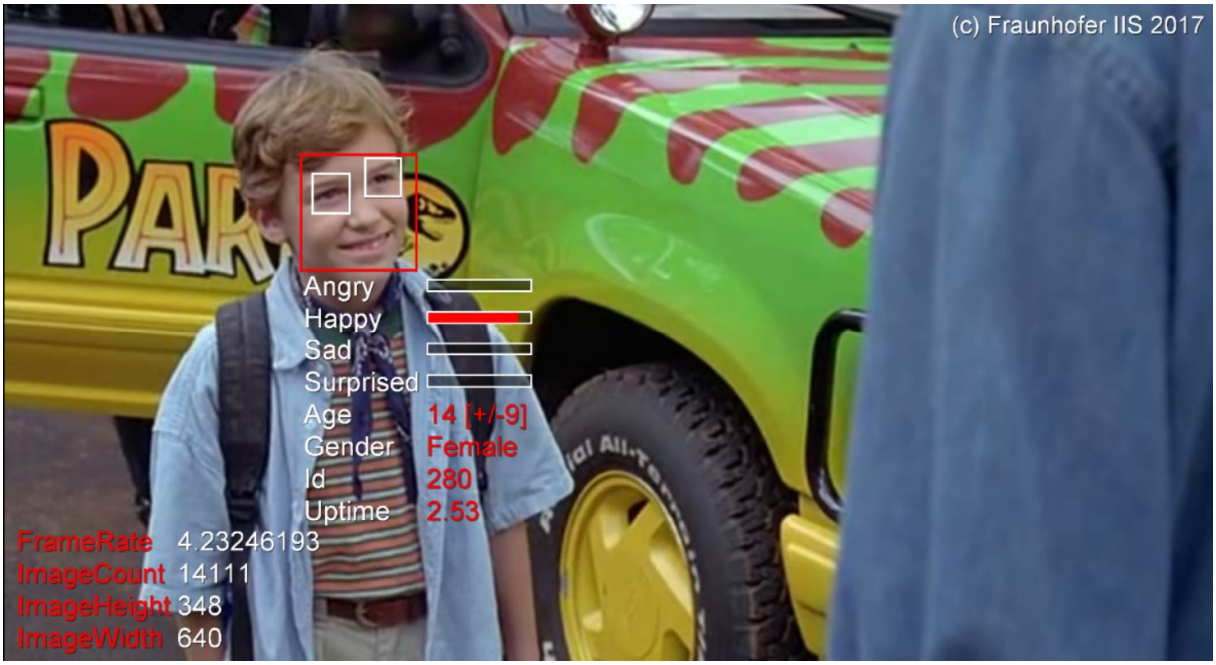
Anhang 21 SHORE-Analyse Dat22



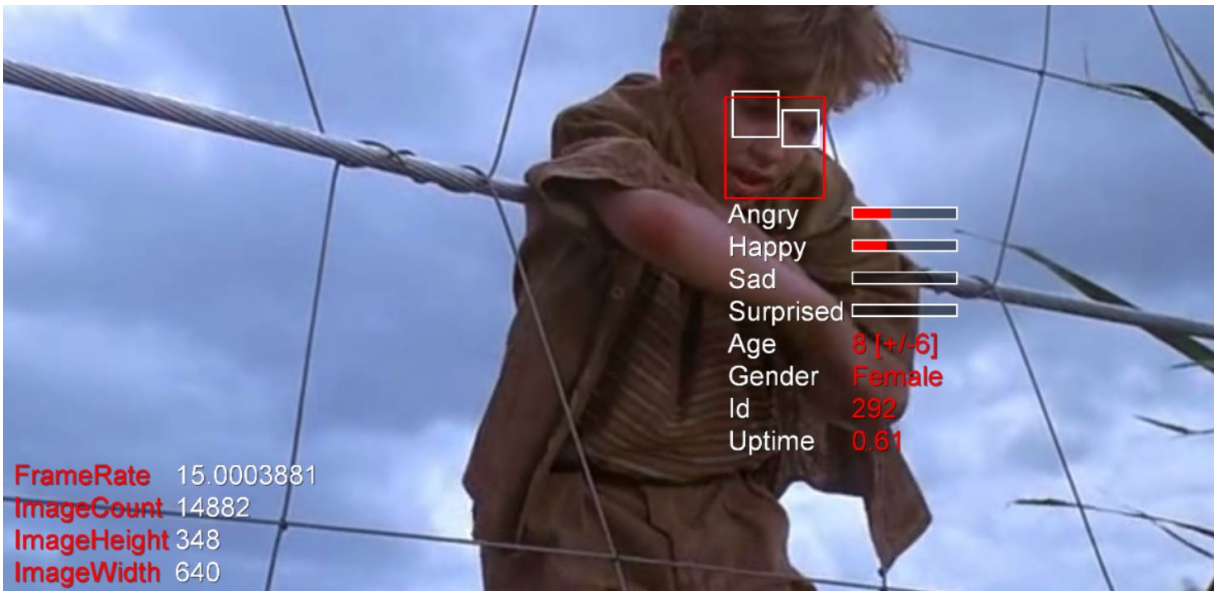
Anhang 22 SHORE-Analyse Dat23



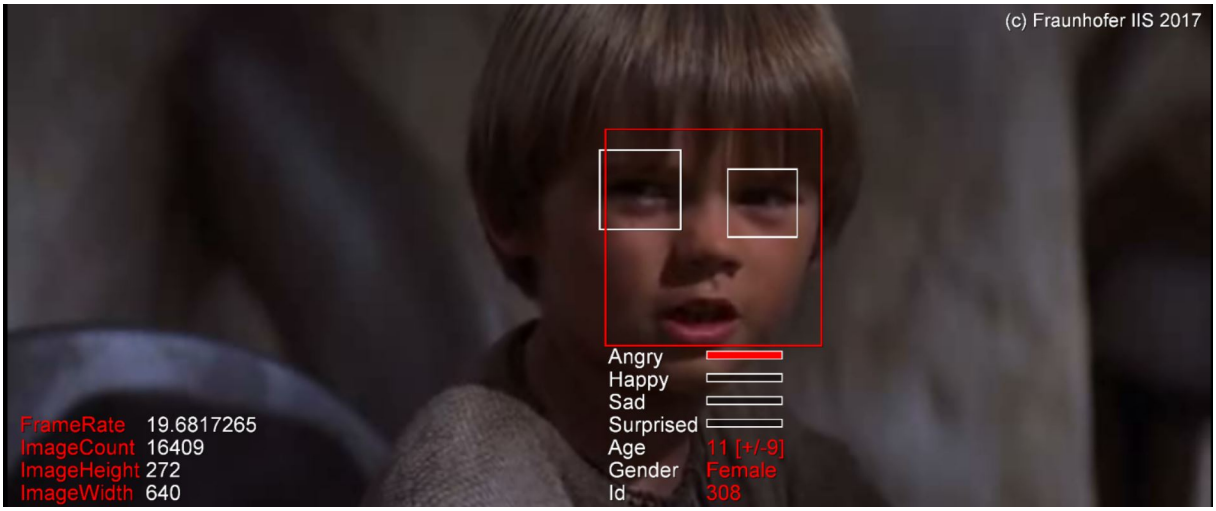
Anhang 23 SHORE-Analyse Dat24



Anhang 24 SHORE-Analyse Dat25

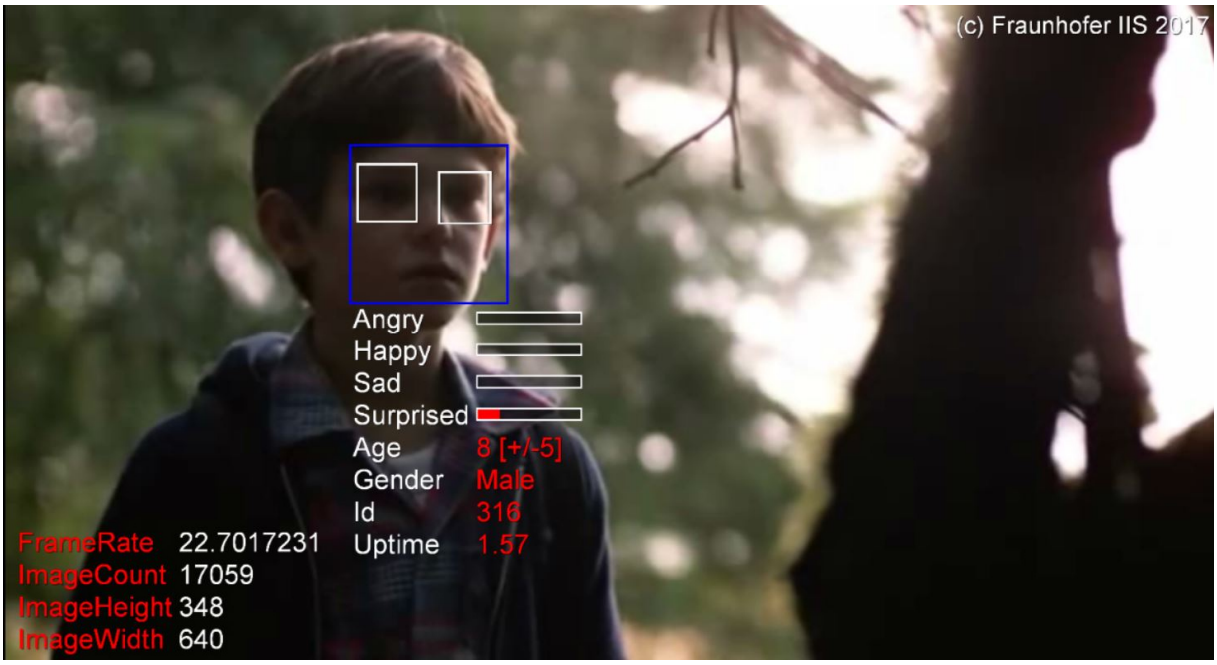


Anhang 25 SHORE-Analyse Dat26

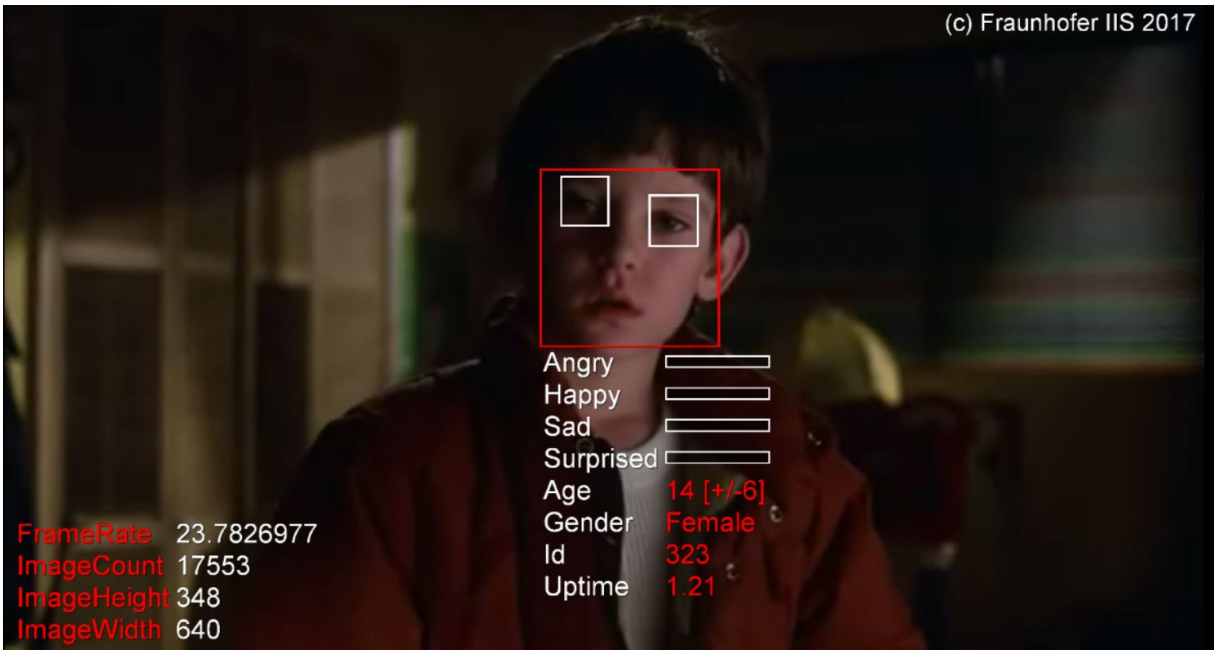


Anhang 26 SHORE-Analyse Dat27

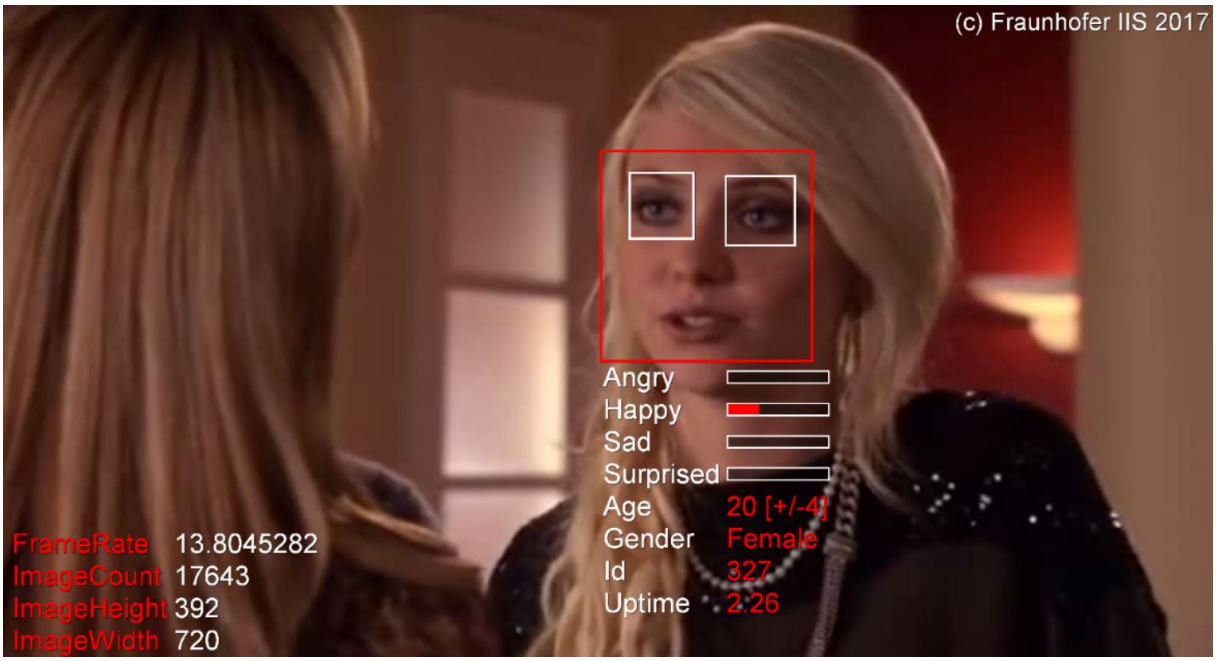
Anhang 27 SHORE-Analyse Dat28



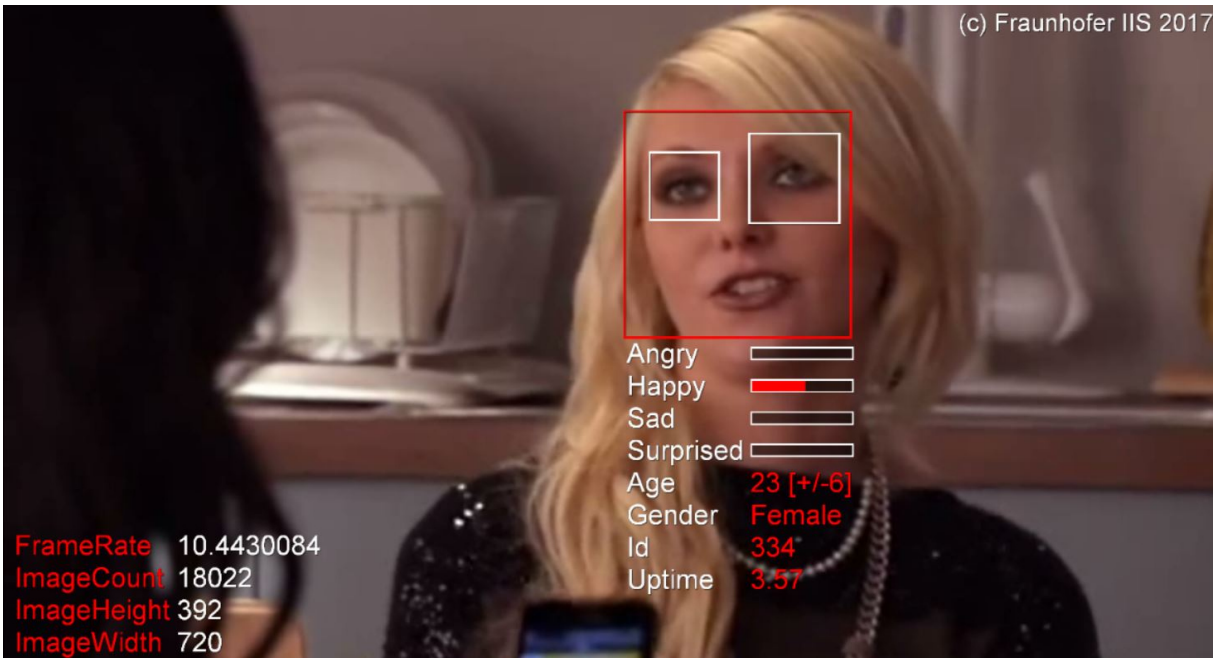
Anhang 28 SHORE-Analyse Dat29



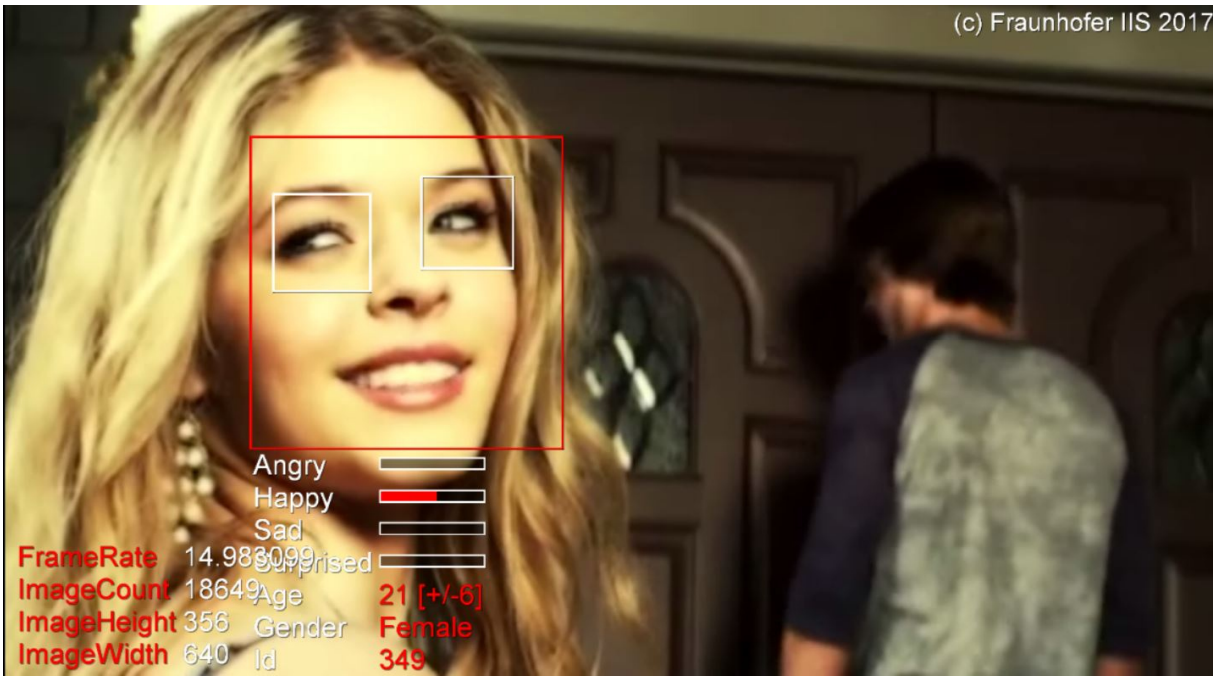
Anhang 29 SHORE-Analyse Dat30



Anhang 30 SHORE-Analyse Dat31

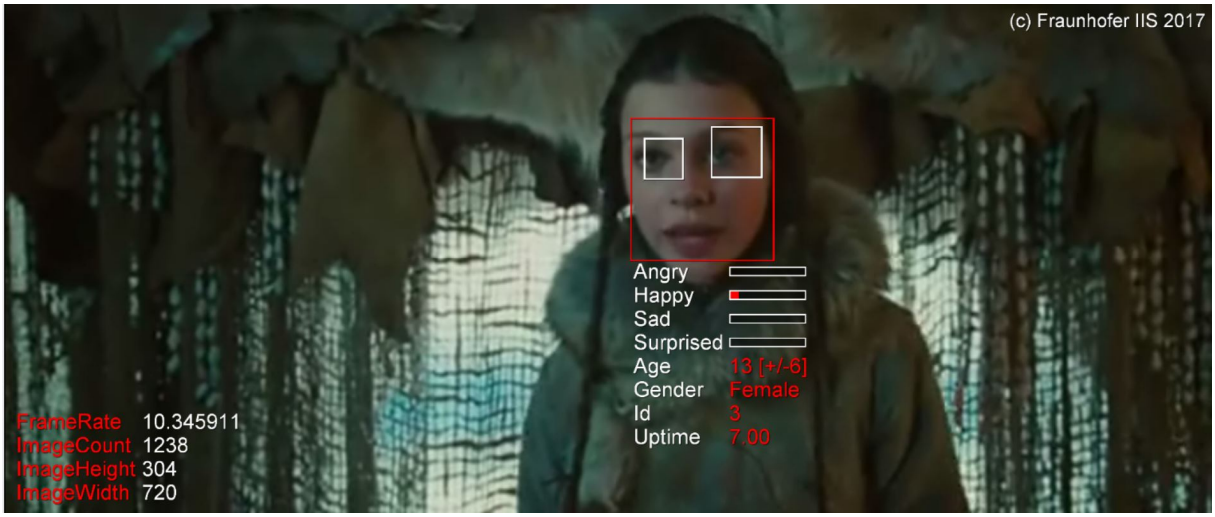


Anhang 31 SHORE-Analyse Dat32

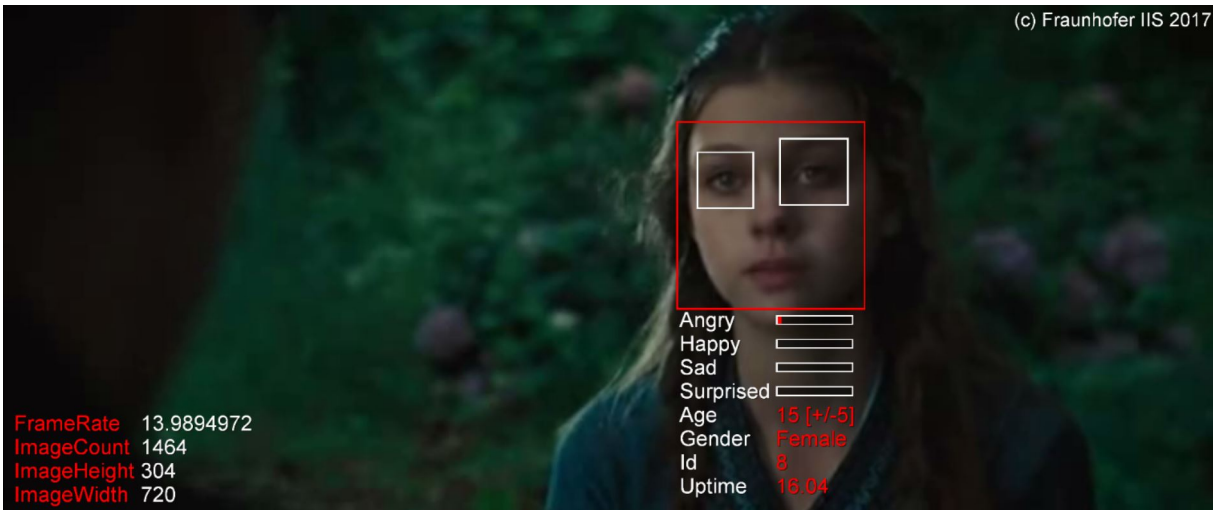


Anhang 32 SHORE-Analyse Dat33

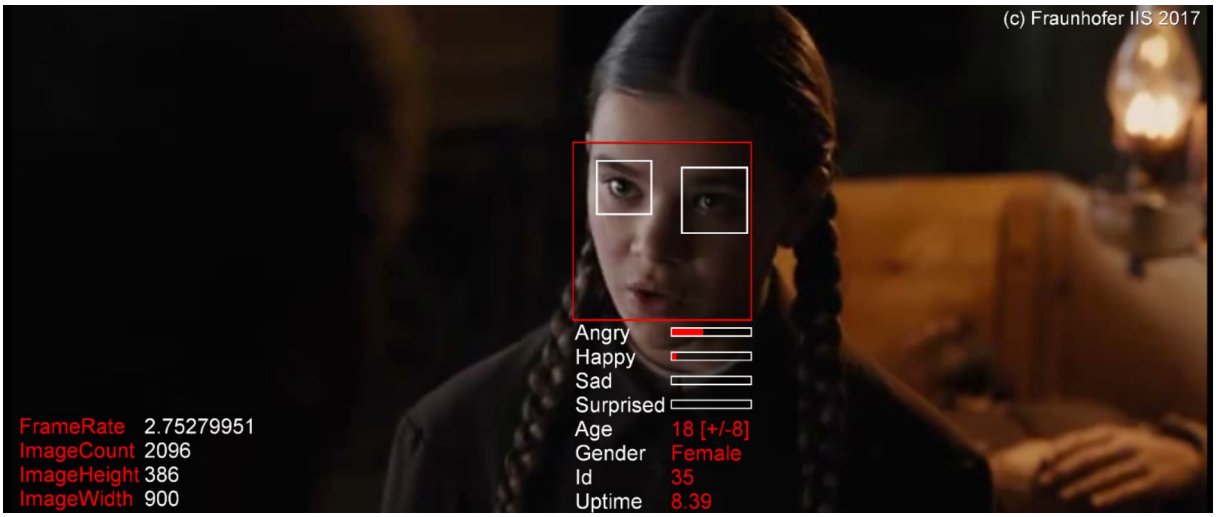
Anhang 33 SHORE-Analyse Dat34



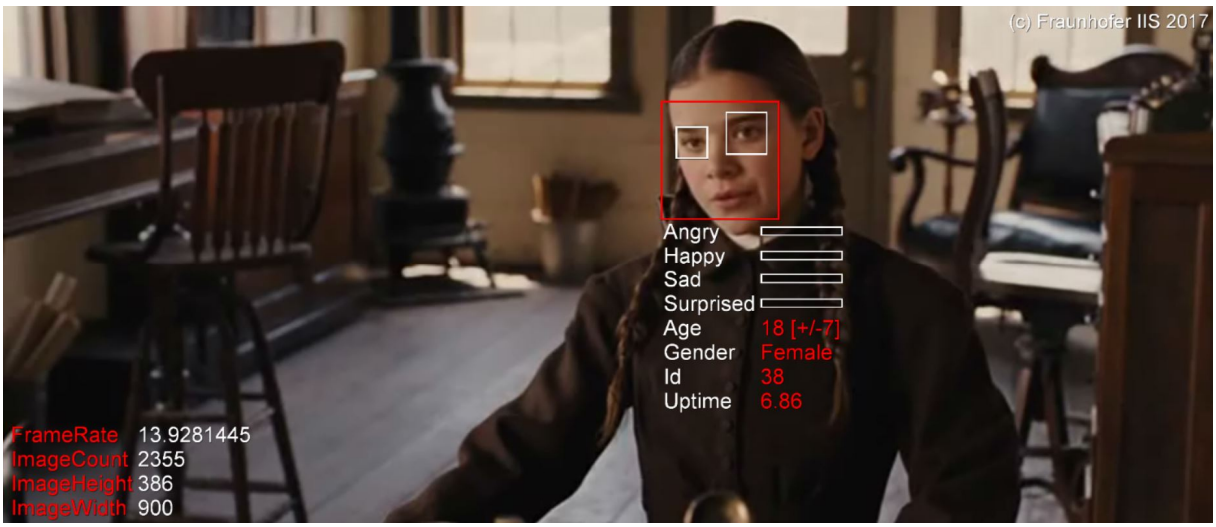
Anhang 34 SHORE-Analyse Dat35



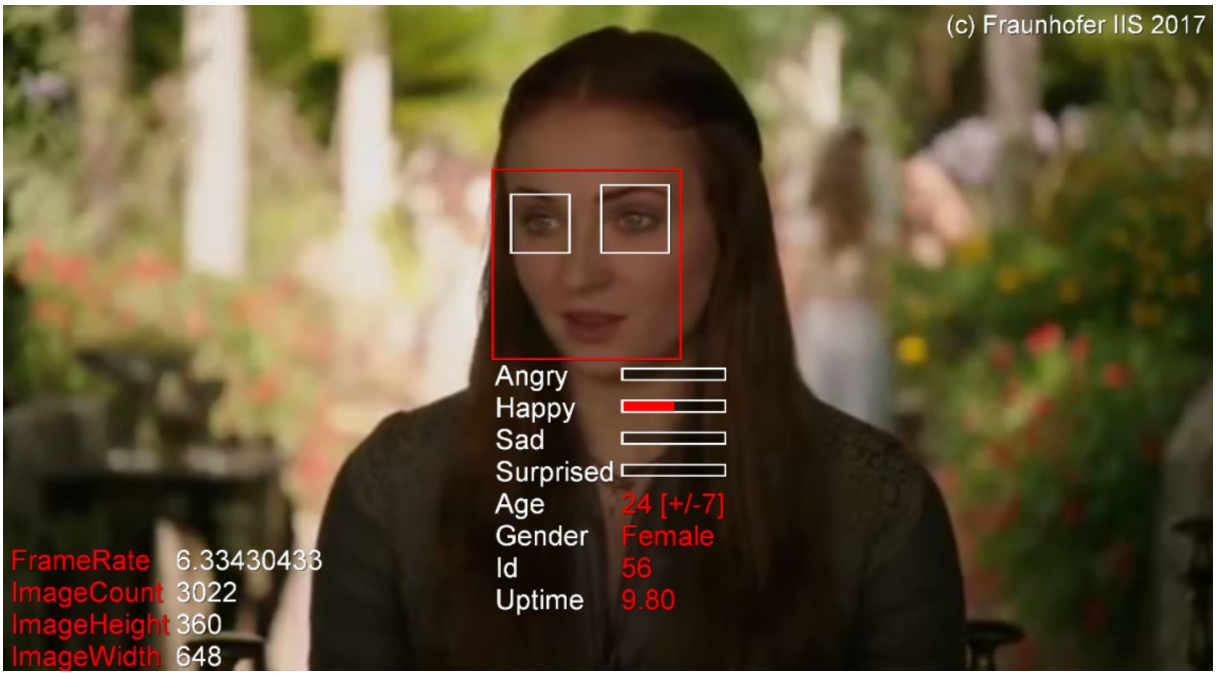
Anhang 35 SHORE-Analyse Dat36



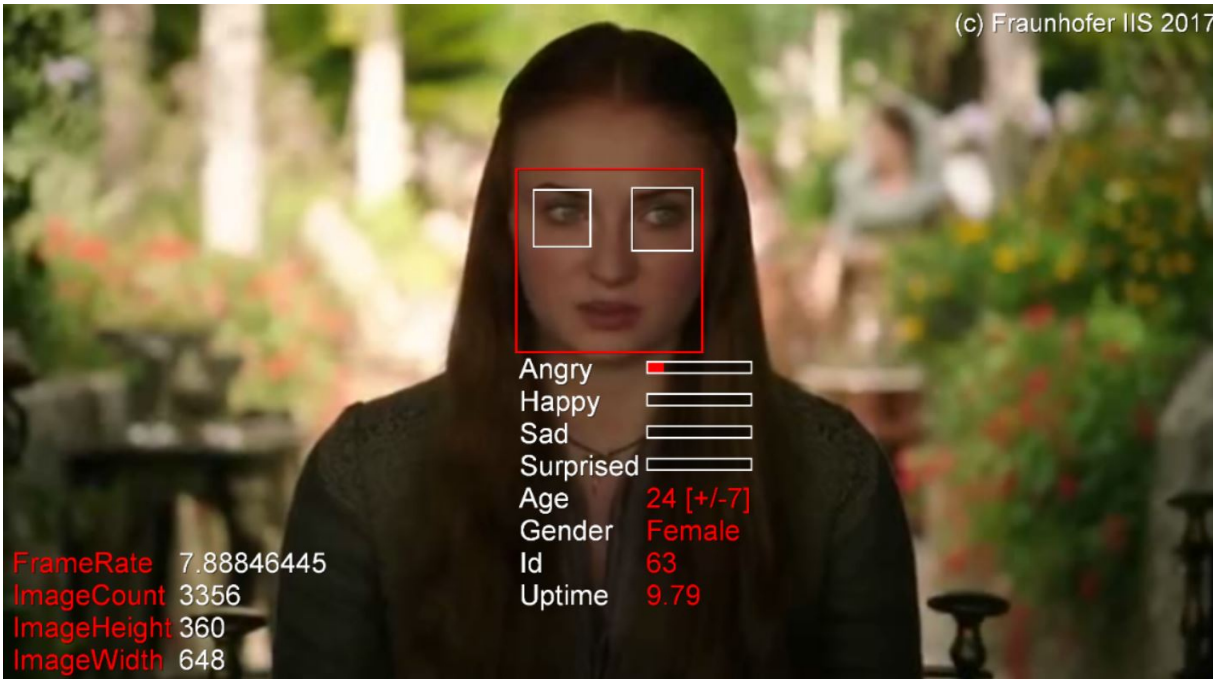
Anhang 36 SHORE-Analyse Dat37



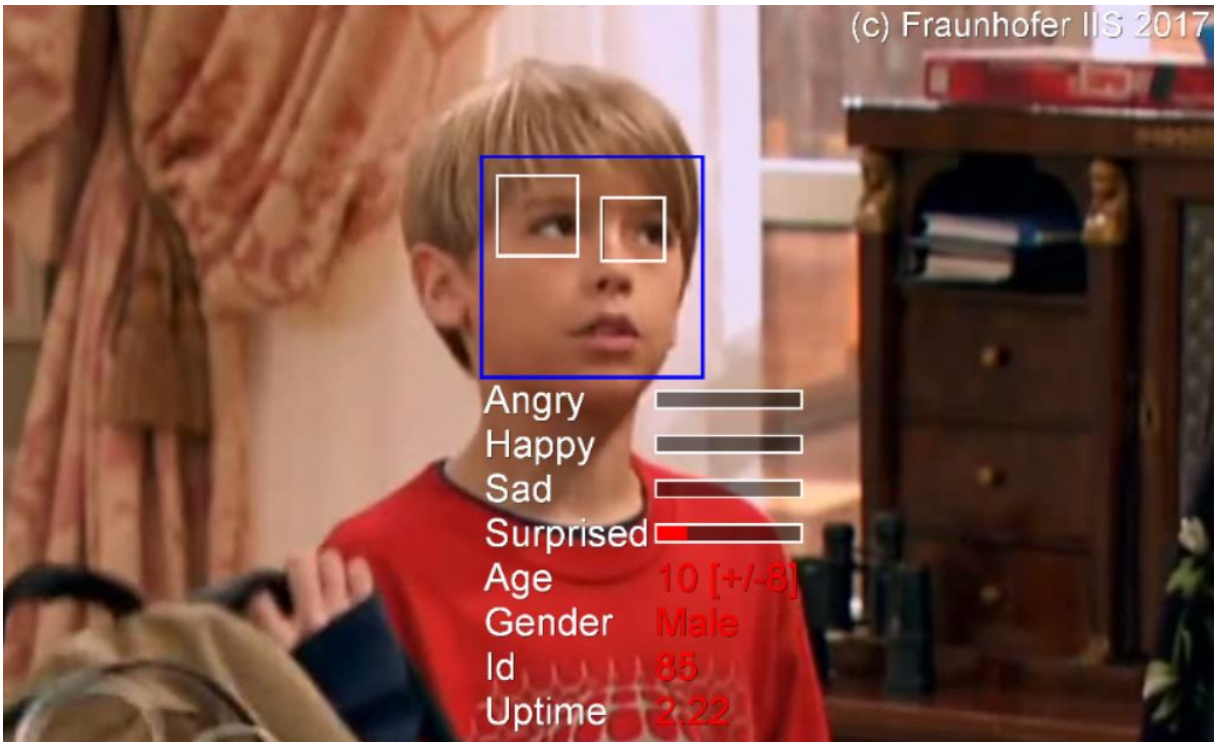
Anhang 37 SHORE-Analyse Dat38



Anhang 38 SHORE-Analyse Dat39



Anhang 39 SHORE-Analyse Dat40

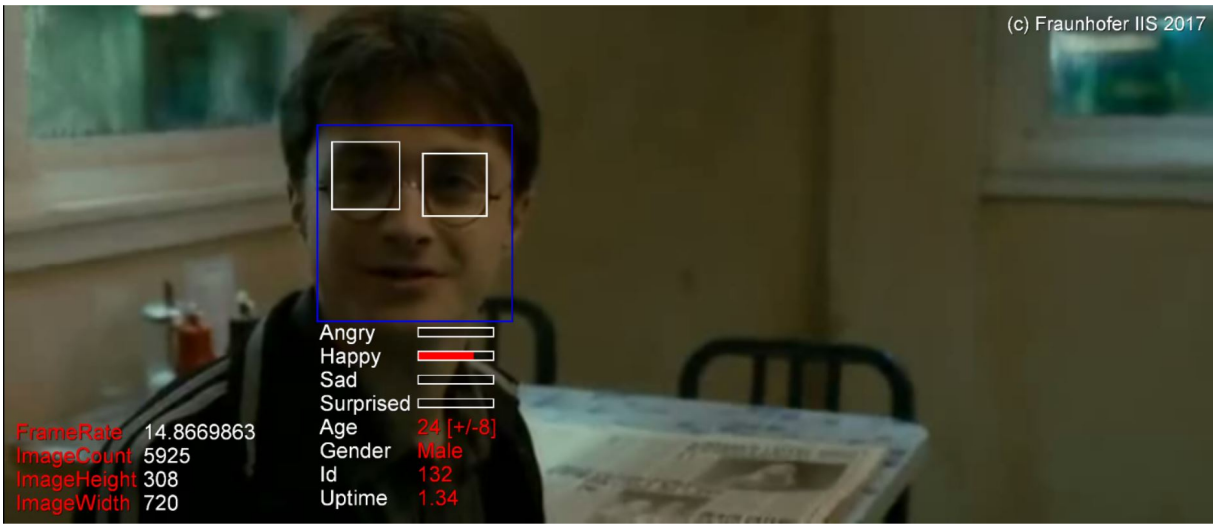


Anhang 40 SHORE-Analyse Dat41

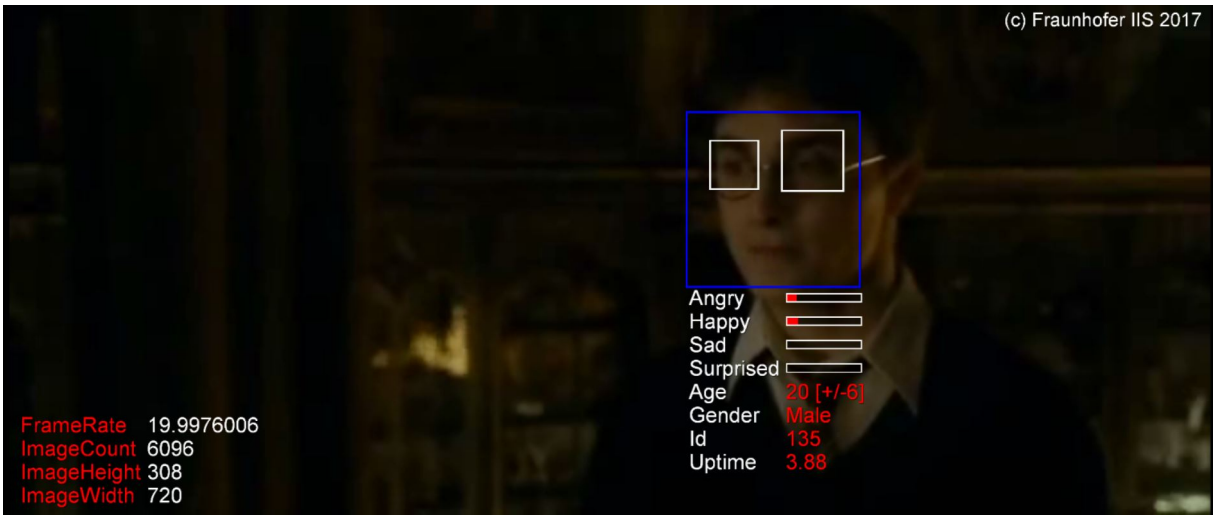
Anhang 41 SHORE-Analyse Dat42

Anhang 42 SHORE-Analyse Dat43

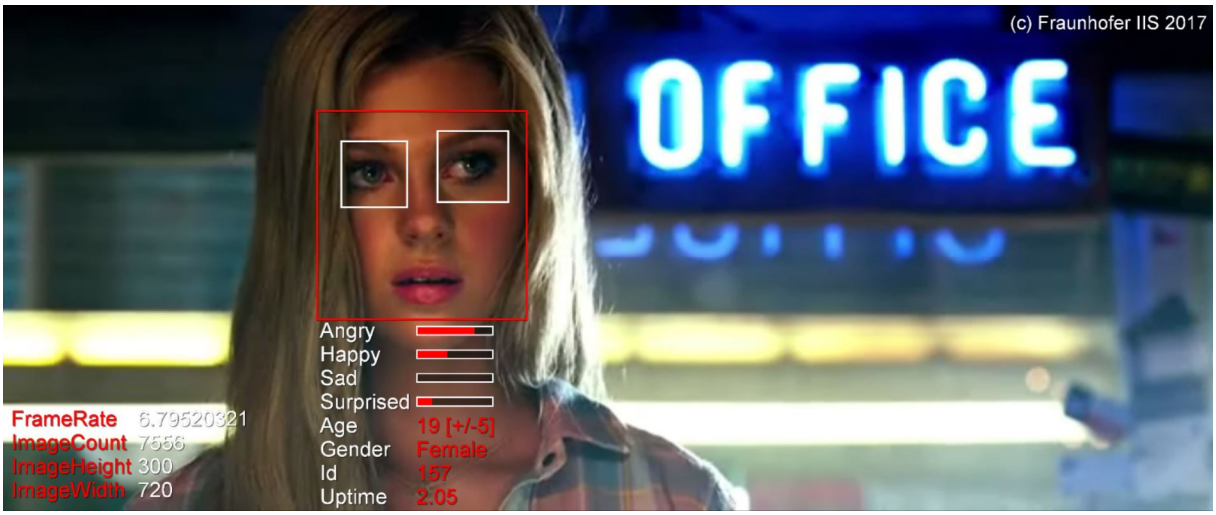
Anhang 43 SHORE-Analyse Dat44



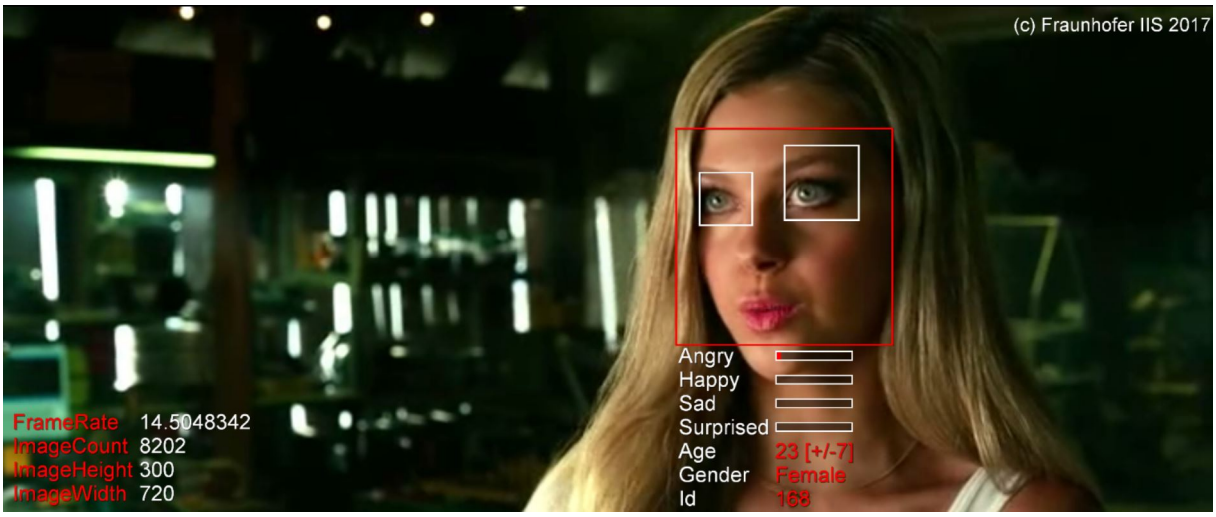
Anhang 44 SHORE-Analyse Dat45



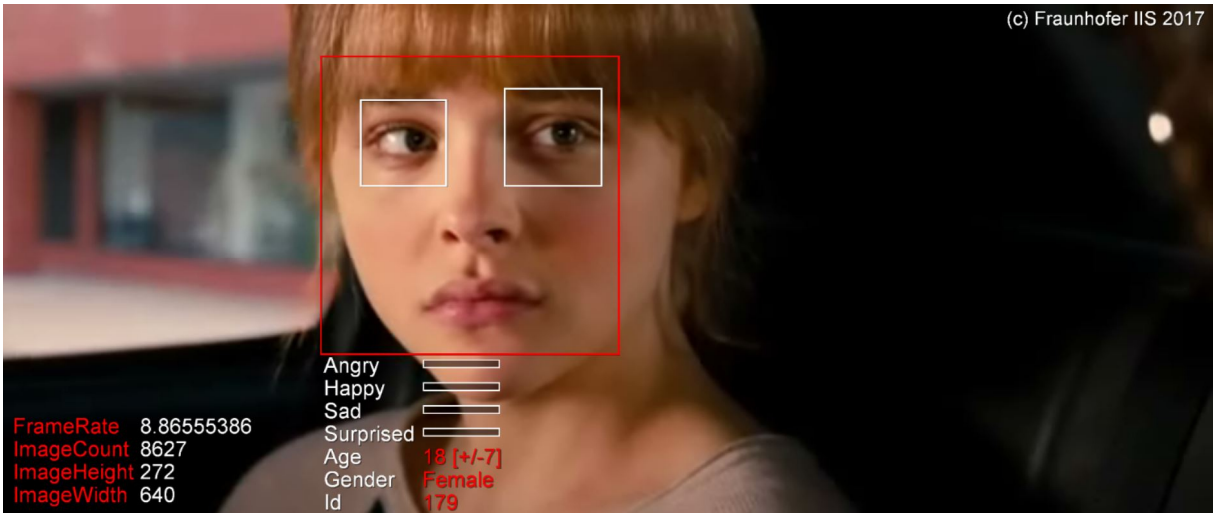
Anhang 45 SHORE-Analyse Dat46



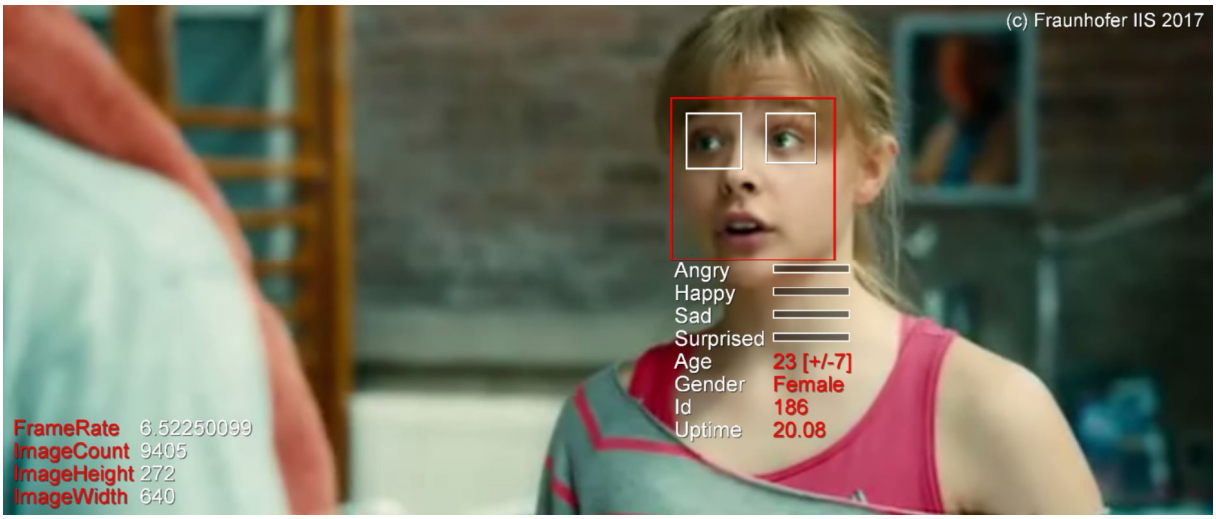
Anhang 46 SHORE-Analyse Dat47



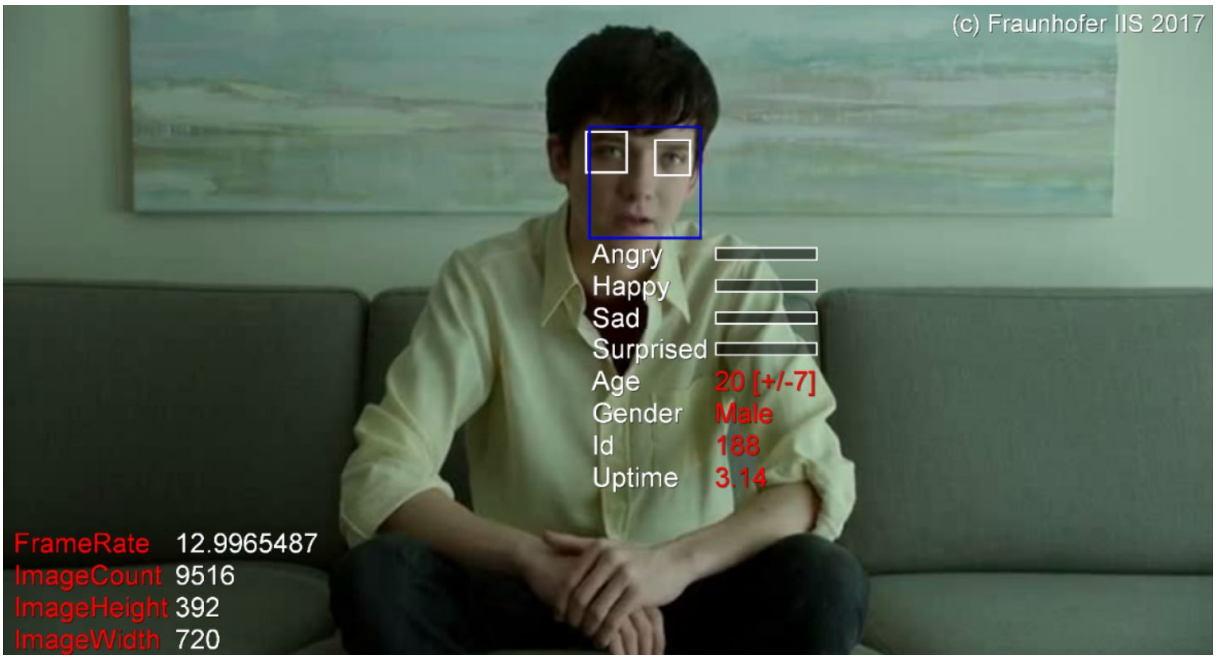
Anhang 47 SHORE-Analyse Dat48



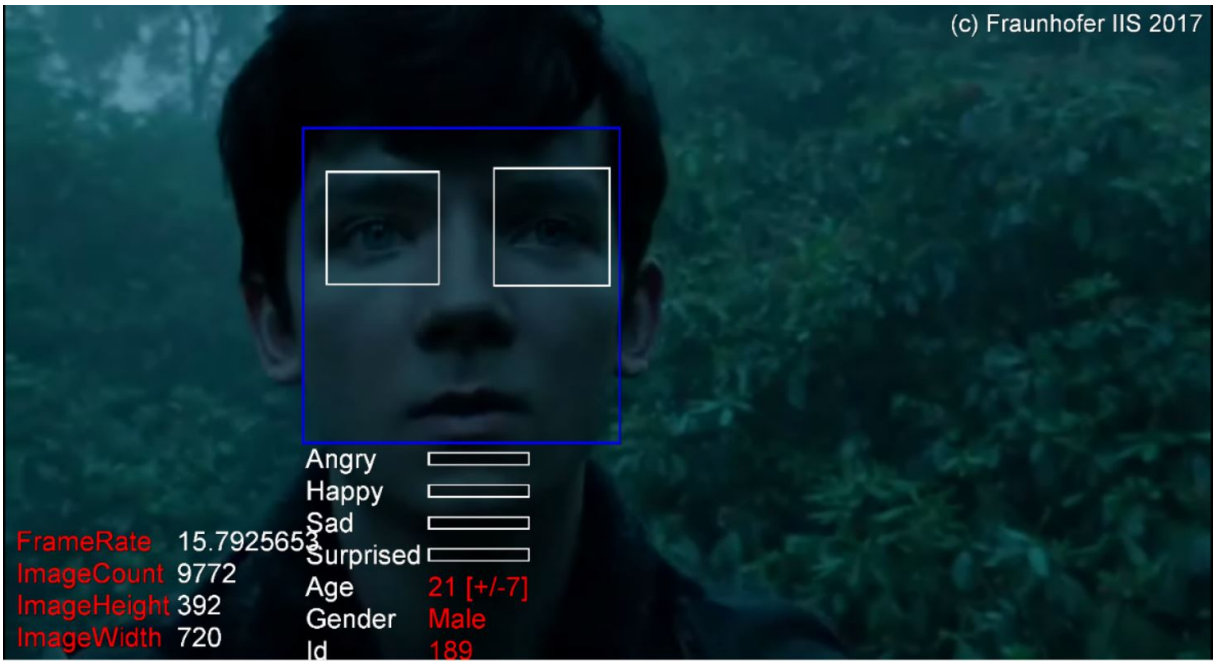
Anhang 48 SHORE-Analyse Dat49



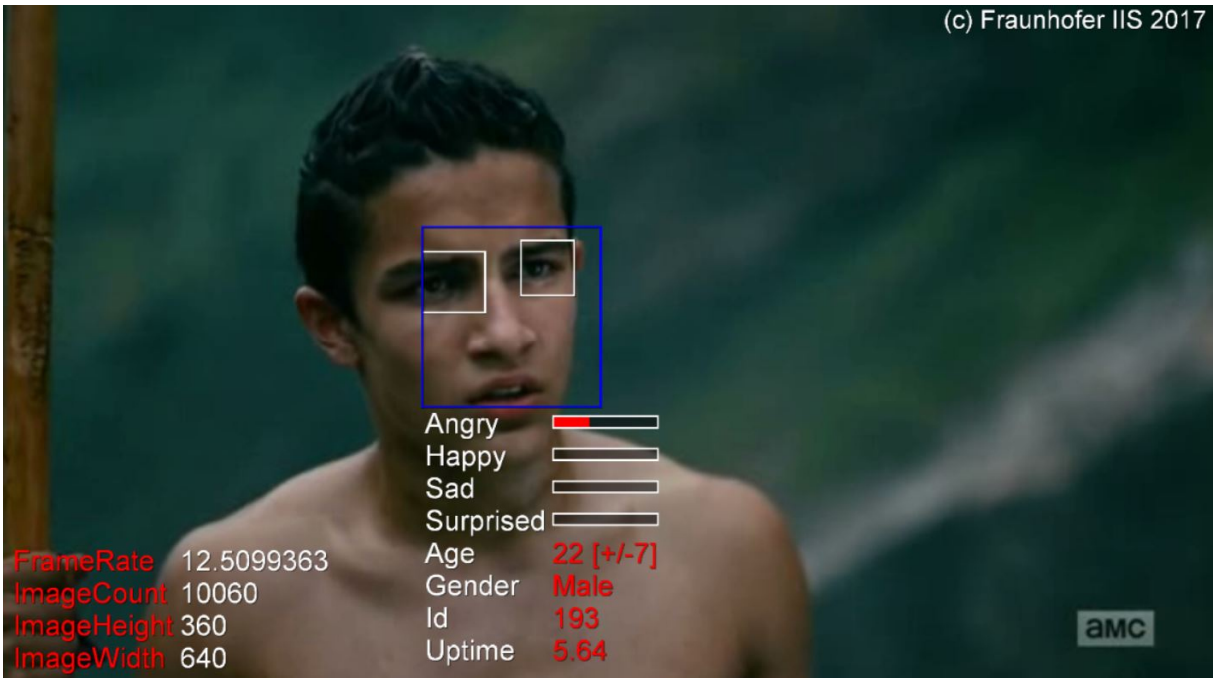
Anhang 49 SHORE-Analyse Dat50



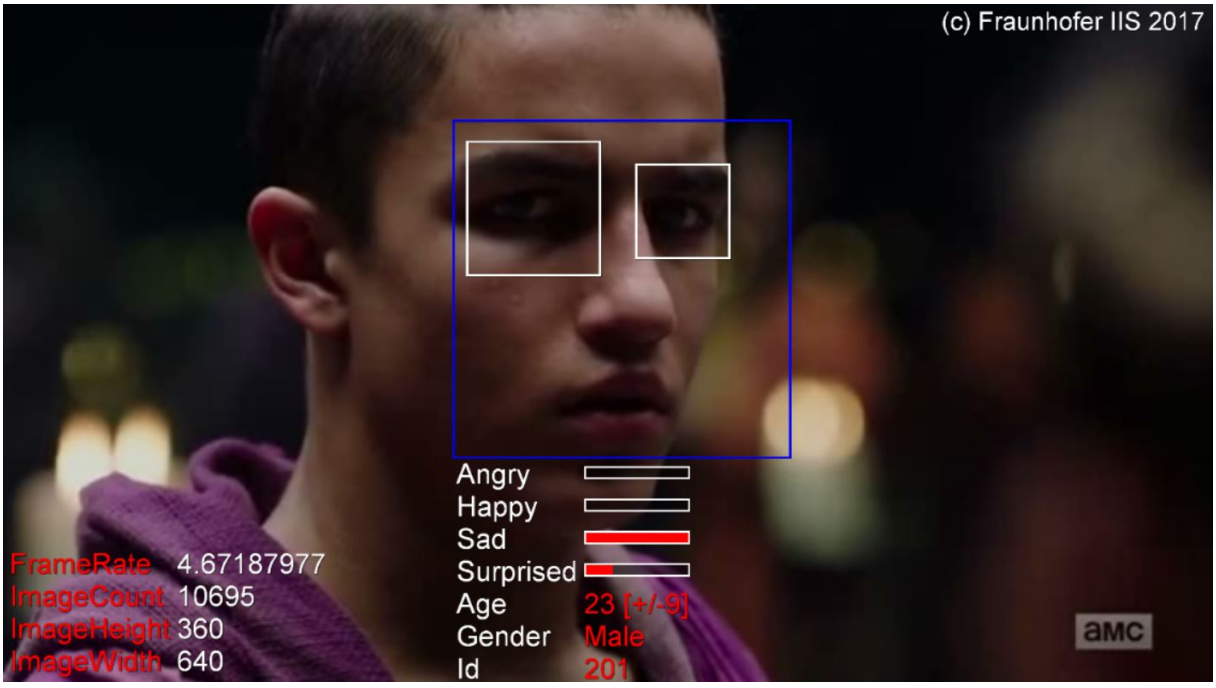
Anhang 50 SHORE-Analyse Dat51



Anhang 51 SHORE-Analyse Dat52



Anhang 52 SHORE-Analyse Dat53



Anhang 53 SHORE-Analyse Dat54

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Leipzig, den 07. August 2017

Christian Klaus